University of Alberta

Characterization of the Zebrafish Immunoglobulin Heavy Chain E(mu)3′ Enhancer

by

Kristofor Kenneth Ellestad

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

in

Physiology and Cell Biology

Department of Biological Sciences

Edmonton, Alberta

Fall 2005

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

0-494-09158-4

Your file *Votre référence*
ISBN:
Our file *Notre reference*
ISBN:

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract:

All examined tetrapod immunoglobulin heavy chain (*IgH*) loci feature a transcriptional enhancer, Eμ, which is located in the J_H-Cμ1 intron and required for *IgH* expression and diversity-generating germline modifications occurring at the locus. Studies of the channel catfish *IgH* locus however demonstrated only the presence of an enhancer in the Cμ-Cδ intergenic region, called Eμ3'. This enhancer position is predicted to be unfavorable for affinity maturation in fish, and would have been a constraint to the evolution of class switch recombination as it developed in the tetrapods. To determine if Eμ3' is a general feature of teleosts, we analyzed the zebrafish *IgH* locus and found an Eμ3'-like enhancer, which shares with catfish Eμ3' a core region of homology and a closely spaced pair of E-box motifs. These E-boxes are critical to enhancer function. The presence of similarly spaced E-boxes in the predicted Eμ3' regions of other fish species indicate that they may be a common feature of fish *IgH* enhancers.

ACKNOWLEDGEMENT

First and foremost, I would like to thank Dr. Brad Magor for giving me the opportunity to undertake this research in his laboratory, as well as his mentorship, support, and guidance throughout the course of my studies. I would also like to thank the members of the labs of Dr. Kathy Magor and Dr. Mike Belosevic – the other two immunology labs in our department - for our collaborations and helping to provide an enjoyable and stimulating environment in which to work.

Thank you to the Natural Sciences and Engineering Research Council of Canada for funding our scientific endeavors and those of so many other investigators in this country.

Finally, to my family, who have stood by my side over the years and supported me in more ways than can be expressed here – I dedicate this thesis to you.

TABLE OF CONTENTS

Introduction....1

Immunoglobulin structure....1

Immunoglobulin genes....4

B-cell activation, affinity maturation, and the evolution of *IgH* loci....7

Promoter, enhancers, and transcription factors control gene expression....18

The function of transcriptional regulatory elements in the *IgH* locus....21

Basic structure of the Eμ enhancer....23

Elements flanking the minimal Eμ enhancer region that contribute to transcriptional activation....29

The immunoglobulin heavy chain enhancer of the Channel Catfish – Eμ3'....34

Structure of the Catfish Eμ3' enhancer....34

Implications of the location and structure of Catfish Eμ3'....36

Hypothesis and objectives....38

Materials and Methods....40

Cloning and sequencing of regions of the Zebrafish *IgH* locus....40

Reporter plasmids....41

E-box mutagenesis constructs....45

Growth and transfection of cell lines....45

Reporter assays....46

DNA sequencing....48

In silico analyses of enhancer regions....48

Results....50

The Zebrafish *IgH* locus contains a functional Eμ3' enhancer....50

The Zebrafish and Catfish Cμ-Cδ intergenic regions share a similar enhancer organization....55

E-boxes are key elements in the function of the zebrafish Eμ3' enhancer....61

Elements flanking the zebrafish Eμ3' core region contribute to enhancer function....65

The zebrafish Eμ3' enhancer can drive transcription in murine B-cells....68

The zebrafish Eμ3' is B-cell specific....68

The zebrafish Cμ-Cδ intron shows no evidence of a remnant transposon....72

Discussion....73

Eμ3' enhancers among teleost fish share common functional and structural features....73

Teleost transcription factors implicated in Eμ3' enhancer function....75

Implications of an Eμ3'-like enhancer as a common feature of teleost fish *IgH* loci....78

A putative transcriptional enhancer in the zebrafish J_H-Cζ intron....80

An alternate scenario of *IgH* enhancer evolution....81

Implications of a J_H-Cζ enhancer for affinity maturation in fish....82

Bibliography....84

Appendices....100

Appendix 1 – Chart of compiled reporter activities displayed with 95% confidence intervals....100

Appendix 2 – *IgH* enhancer motifs from the TFD database (Ghosh 1990) included in motif searches 101

Appendix 3 – Raw transfection data 102

Appendix 4 – Appendix 4 – Table of selected high-scoring tBLASTn hits against a putative *HEB* homologue in the zebrafish genome database and multiple alignment of its predicted amino acid sequence with catfish CFEB-1 110

Appendix 5 – Table of selected high-scoring tBLASTn hits against a putative *Oct2* homologue in the zebrafish genome database and multiple alignment of its predicted amino acid sequence with catfish Oct2β 111

LIST OF TABLES

Table 1 - Primers used for generating Zebrafish *IgH* enhancer constructs.....43

Table 2 - Potential transcription factor binding sites in the Zebrafish Eμ3' enhancer .51

Table 3 - Potential transcription factor binding sites in the Zebrafish J_H-Cμ1 intron ..52

LIST OF FIGURES

Figure 1 - Schematic of the structure of an immunoglobulin monomer.....3

Figure 2 - Schematic of the germline murine *IgH* locus before and after VDJ recombination.....5

Figure 3 - Immunoglobulin isotypes expressed among vertebrate lineages.....9

Figure 4 - Schematic of the murine *IgH* locus before and after class switching.....12

Figure 5 - Schematic of the channel catfish and zebrafish *IgH* loci.....14

Figure 6 - Diagrams of the mouse Eμ and channel catfish Eμ3' enhancers.....24

Figure 7 - The relative positions of regions of the zebrafish CμTM1-Cδ intergenic region included in reporter constructs.....44

Figure 8 - Transcriptional enhancer activity from regions of the zebrafish immunoglobulin heavy chain (*IgH*) locus tested in catfish B-cells.....53

Figure 9 - Localizing transcriptional activity within the Cμ-Cδ intergenic region.....56

Figure 10 - Dot-matrix comparisons of fish and mammalian *IgH* enhancer regions...58

Figure 11 - Alignment of homologous regions the zebrafish and catfish Eμ3'.....60

Figure 12 - A pair of conserved E-box motifs are important in mediating transcriptional activity from the zebrafish Eμ3' enhancer.....63

Figure 13 - Scaled maps of the catfish and zebrafish Eμ3' enhancers showing the relative positions of predicted enhancer motifs.....67

Figure 14 - The zebrafish Eμ3' enhancer is functional in murine B-cells.....69

Figure 15 - The zebrafish Eμ3' enhancer is B-cell specific.....71

LIST OF ABBREVIATIONS

AID	activation induced cytidine deaminase
ARID	A-T rich interacting domain
b -	basic
BCR-	B-cell receptor
BLAST	Basic local alignment search tool
Bright	B-cell regulator of *IgH* transcription
C-	constant
CAT-	chloramphenicol acetyltransferase
CBF	core binding factor
CSR	class switch recombination
D-	diversity
DNA	deoxyribonucleic acid
Eμ-	*IgH* intronic enhancer
GC	gene conversion
H-	heavy chain
HLH	helix-loop-helix
IgH	immunoglobulin heavy chain
J-	joining
Ig-	immunoglobulin
IRF	interferon regulatory factor
MAR	matrix attachment region
Mya	millions of years ago
OBF-1	octamer transcription factor binding factor
Oct-	octamer transcription factor
PAC	P1 artificial chromosome
PCR	polymerase chain reaction
RNA	ribonucleic acid
RT-	reverse transcriptase
S-	switch region (CSR)
SH	somatic hypermutation
ssDNA	single stranded DNA
TM	transmembrane
V-	variable
ZIP	leucine zipper

Introduction:

In order to be able to recognize and eliminate potential pathogens, the vertebrate adaptive immune system relies on antigen receptor genes which have evolved the potential to recombine in the germline to encode receptors with diverse ligand binding specificities. During the development of B-cells a number of germline modifications to the immunoglobulin genes allow each B-cell to encode antibodies of a unique specificity. If a B-cell encounters a pathogen to which its antibody can bind and becomes activated, the immunoglobulin genes in that B-cell can undergo further modifications that increase antibody binding affinity and direct secondary responses appropriate to the location of infection and the nature of the immune challenge. Appropriate gene expression and all of these germline modifications depend on the tightly regulated control of transcription, which is mediated by cell type specific promoters and enhancers. This thesis examines the location and structure of the immunoglobulin heavy chain gene enhancer(s) in a teleost fish, with a view to understanding how enhancer organization and location influenced the elaboration of humoral immunity over the course of vertebrate evolution.

Immunoglobulin structure

Immunoglobulins (Ig), or antibodies, are a key component of the adaptive immune response of the jawed vertebrates. They are present in serum as a soluble form where they can bind to molecular components of pathogens and facilitate their removal or inactivation. They can also be present as a membrane bound form on the surface of B-cells in association with integral membrane proteins in a structure known as the B-cell receptor (BCR). The basic immunoglobulin subunit is a tetrameric molecule composed

of two identical heavy and two identical light chain peptides, linked together with disulfide bonds (Fig. 1). The heavy and light chain peptides are encoded by different genes. Both the light chain and heavy chain peptides are composed of a repeating substructure known as an immunoglobulin domain (reviewed in Frazer and Capra 1999). Light chains are composed of one variable (V_L) and one constant (C_L) domain, while heavy chains are composed of a single V_H and multiple (two or more) C_H domains. These domains all share a common structural motif known as the immunoglobulin fold, which is comprised of β-strands organized into two disulfide bond joined β-pleated sheets. The β-strands are joined by loops of variable length in order to facilitate β-sheet formation.

The variable region of the antibody, which is involved in ligand binding and is unique to each antibody-expressing B-cell, is comprised of the V_H and V_L domains, which are hypervariable with respect to the amino acid sequence of portions of the loops connecting the β-strands. These hypervariable regions are known as the complementarity determining regions (CDRs), and participate in formation of the ligand binding site. Each immunoglobulin monomer has two such identical binding sites.

The constant, or Fc region of the antibody, has numerous biological effector functions and is comprised of the relatively non-variable C_H domains of the heavy chain peptides (Fig. 1). The identity of the constant region of the antibody heavy chain determines its isotype (eg: IgM, IgD, IgG, etc.), and in many vertebrates B-cells can vary their expressed Fc region after activation by encounter with antigen and the appropriate co-stimulatory signals. The effector functions mediated by a particular antibody isotype, for example docking with Fc receptors or fixing complement, can vary. Depending on

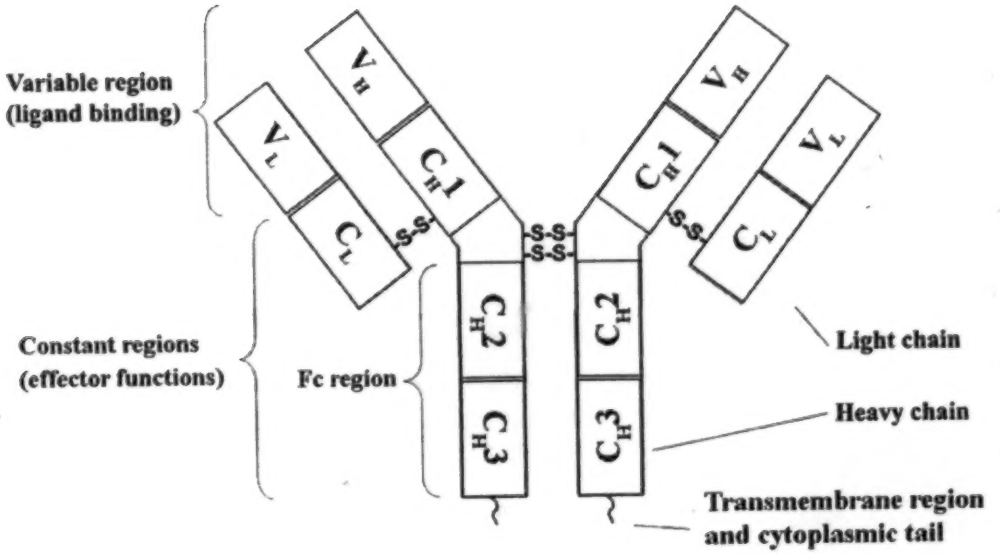


Figure 1 – Simplified schematic of the structure of an immunoglobulin monomer. Each immunoglobulin monomer is comprised of two heavy and two light chain peptides. These peptides are held together by intermolecular disulfide bonds. Each of the heavy and light chain peptides are composed of one variable (V) immunoglobulin domain and one or more constant (C) immunoglobulin domains. The V domains of the heavy (V_H) and light (V_L) chains comprise the variable region or ligand binding portion of the immunoglobulin monomer. The constant region of the antibody, which has numerous biological effector functions, is comprised of the C_L and C_H domains. Immunoglobulins can be present as a secreted or membrane-bound form. In the membrane-bound form, alternate mRNA splicing of the *IgH* transcript gives rise to a C-terminal transmembrane region on the heavy chain peptide.

characteristics of their isotype, immunoglobulin monomers may also further multimerize. For example, in humans, IgM is present in serum as a pentamer. In teleost fish, IgM in serum forms a tetramer. The number of expressed antibody isotypes varies among the vertebrate lineages and has undergone elaboration over the course of evolution. The genetic basis of these differences lies in variations in the structure of their immunoglobulin heavy chain genes, which will be discussed below.

Immunoglobulin genes

Diversity in the antigen binding portion of the immunoglobulin molecule is the key to effective humoral immunity, and among the jawed vertebrates several genetic mechanisms have evolved to achieve it. These processes either act on the immunoglobulin gene during initial B-cell development, or following activation of the B-cells by encounter with their specific antigen. VDJ recombination occurs in all jawed vertebrates. Lineages diverged after the cartilaginous fish, which include amphibians, reptiles, birds, mammals, and teleost fish, share the "translocon" (or V_{1-n}-D_{1-n}-J_{1-n}-C) arrangement of their immunoglobulin heavy chain (*IgH*) genes (Fig. 2a) (reviewed in Flajnik 2002). In these organisms many gene segments (or sub-exons) termed V, D, and J are clustered in the 5' end of the germline *IgH* locus upstream of an array of constant region exon clusters or C elements. A VDJ recombinase complex serves to "randomly" join single V, D, and J elements together to form a contiguous VDJ exon encoding the variable region of the antibody (Fig. 2b) (reviewed in Max 1999). This process involves the excision of intervening regions of somatic germline DNA, bringing together members

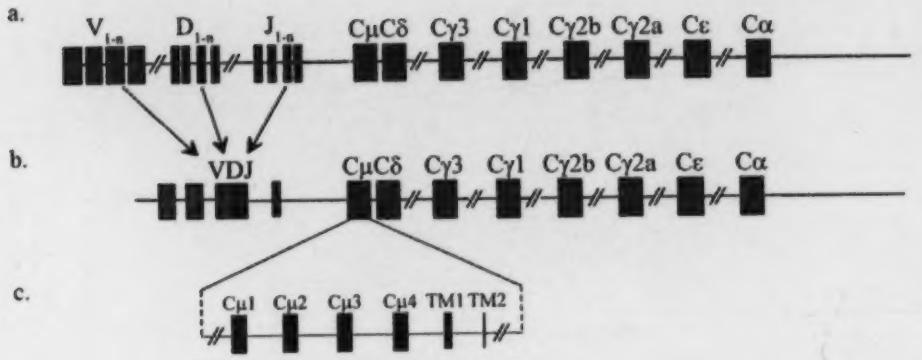


Figure 2 – Schematic (not to scale) of the murine *IgH* locus before (a) and after VDJ recombination (b). The translocon arrangement of multiple V, D, and J elements or sub-exons that recombine to form a continuous VDJ exon encoding a functional Ig heavy chain variable region is conserved among the *IgH* loci of the bony fish and later diverged vertebrate lineages. Downstream of the V, D, and J elements are a series of constant (C) region elements, each of which is composed of multiple exons (c) including transmembrane (TM) exons that encode the hydrophobic C-terminal tail of membrane-bound immunoglobulin. The most VDJ proximal C element encodes the constant region of the expressed heavy chain peptide. After B-cell activation, class switch recombination serves to bring a particular C element into proximity with the rearranged VDJ exon and thus alters the expressed constant region.

of diversified families of these elements present within the *IgH* locus. Imprecise joining of the gene segments and the introduction of untemplated bases at their junctions during recombination also contribute to the generation of antibody variable region diversity.

Some species such as chickens and rabbits undergo VDJ recombination, but only one or a few V_H elements in the locus are used. These species generate antibody variable region diversity via a process called gene conversion (GC), where V_H pseudogenes in the locus are used to template modifications to the rearranged VDJ exon (reviewed in Arakawa and Buerstedde 2004).

Cartilaginous fish, which are the earliest diverged vertebrates to express immunoglobulin, have many comparatively simpler *IgH* loci which are composed of a few (one or two) V, D, and J elements, upstream of a single C element (reviewed in Litman et al. 2004). These loci undergo VDJ recombination but the potential diversity generated by this process alone is limited. In many species of cartilaginous fish, some or all of these *IgH* loci are already joined in the germline to encode a contiguous VDJ exon.

The VDJ recombination process is preceded by the generation of sterile, or non-coding transcripts through V region genes from promoters upstream of these elements. Production of these transcripts results in chromatin decondensation which provides access for the VDJ recombinase complex (reviewed in Chowdhury and Sen 2004). Following VDJ rearrangement, the newly formed exon is in proximity to the first C element (Cμ in the *IgH* locus of mice and humans (Fig. 2b)) and so the Cμ exon cluster will be transcribed and included in the C-terminal portion of the nascent IgH polypeptide upon translation. Each constant region exon cluster in the *IgH* locus encodes the heavy chain constant region of a particular antibody isotype. C elements are composed of a

variable number of exons, including transmembrane exons that encode the heavy chain carboxy-terminal portion of surface-bound antibody (Fig. 2c). The membrane and secreted forms of a particular constant region arise by alternate splicing of the *IgH* transcript.

B-cell activation, affinity maturation, and the evolution of IgH loci

The *IgH* loci of the jawed vertebrates have undergone considerable elaboration (Fig. 3) over the course of evolution (reviewed in Flajnik 2002). There are variations in the number and types of V, D, J, and C elements within these loci which have arisen by duplication and diversification of a simpler set of elements from a common ancestral locus (reviewed in Eason et al. 2004). These changes have facilitated the evolution or elaboration of mechanisms that act to improve the antibody response following encounter with antigen, both by increasing the diversity of antigens bound and the secondary responses elicited through the constant regions of different antibody isotypes.

The B-cells of amphibians and later diverged vertebrate lineages (tetrapods) have evolved the ability to vary the constant region of the antibody they express through a process known as class switch recombination (CSR), which like VDJ recombination involves excision of a portion of a B-cell's germline *IgH* gene (Fig. 4). Following activation in lymphoid organs of the higher vertebrates such as mice and humans, some proliferating B-cells form germinal centers where they undergo CSR and a process known as somatic hypermutation (SH). Species which undergo gene conversion to develop their primary variable region repertoire can also undergo GC after B-cell activation (reviewed in Flajnik 2002). CSR, SH, and GC each allow refinement of a B-

cell's expressed immunoglobulin gene to improve the response to a particular immune insult. Such refinement is collectively referred to as affinity maturation.

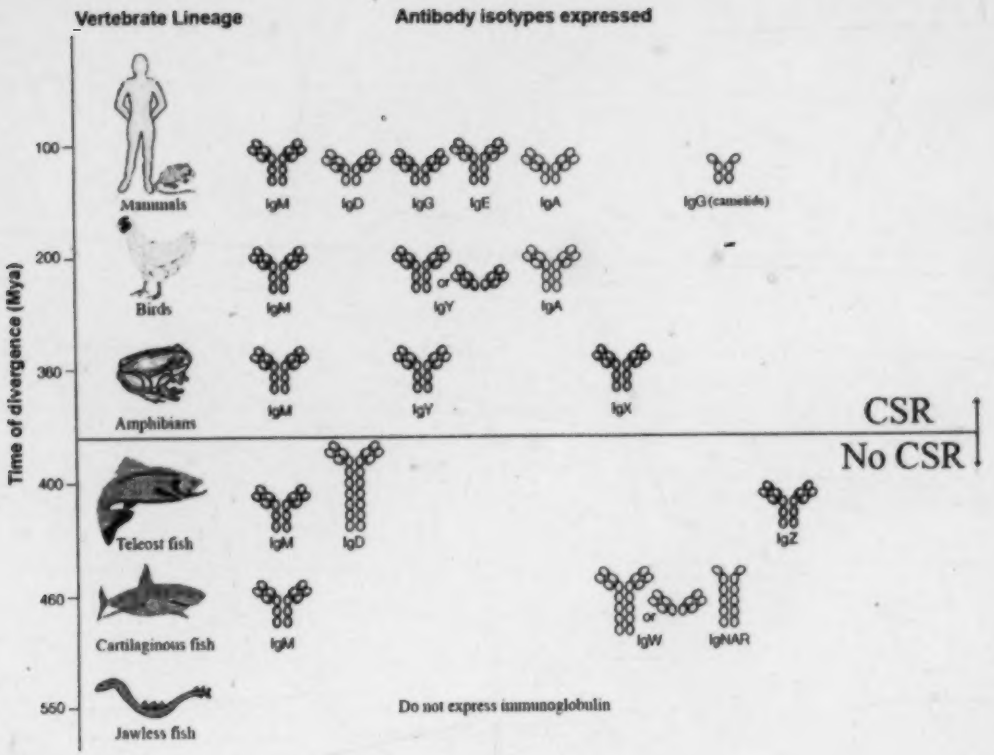


Figure 3 – Immunoglobulin isotypes expressed among vertebrate lineages. Humoral immunity has undergone considerable elaboration with respect to the number and identity of expressed antibody isotypes over the course of vertebrate evolution. In jawless vertebrates no evidence of immunoglobulin expression has been found. Although cartilaginous fish express antibodies, they have many simpler immunoglobulin heavy chain loci which are arranged in a "clustered" configuration where very few (one or two) V, D, or J elements are situated upstream of a C element (Flajnik 2002; Litman et. al. 2004). In many cartilaginous fish species some or all of these immunoglobulin loci are also germline joined and do not undergo VDJ recombination. The bony fish are the earliest diverged vertebrates known to organize their *IgH* locus in the translocon (or $V_{1\text{-}n}$,

$D_{1\text{-}n}$, $J_{1\text{-}n}$, C_n) configuration. The divergence of the tetrapods coincides with the emergence of class switch recombination (CSR) which is an activation induced cytidine deaminase (AID) dependent process. AID is also involved in somatic hypermutation (SH) which occurs in all jawed vertebrates to varying extents. Mya – millions of years ago. Figure adapted from Flajnik, 2002 and Litman et. al. 2004.

The germline *IgH* locus is probably too large for a complete pre-mRNA containing the full array of C elements (~2.5 Mbp from J_H-Cα in the mouse). CSR functions after B-cell activation to bring a particular C-element into proximity with the rearranged VDJ exon such that it will be transcribed and translated as part of the heavy chain (Fig. 4). This process allows the selection of an antibody Fc region for its unique biological effector functions – a method of effectively tailoring the immune response to a particular challenge. A switch (S) sequence is located 5' of each C element and CSR is achieved by recombination between these switch sequences with excision of intervening DNA. The class switch to a particular isotype is dependent on transcription from cryptic promoters upstream of the S regions in concert with enhancers present elsewhere in the locus (reviewed in Chaudhuri and Alt 2004). Transcription through G-C rich switch regions upstream of the constant region genes in mammalian *IgH* loci has been shown to generate stable R-loops *in vivo* (Yu et al. 2003) which are RNA-DNA hybrid structures forming at the transcription bubble where the switch region transcript is hybridized to the template DNA. This results in the formation of single stranded DNA on the nontemplate strand, which represents a substrate for a key component of the CSR machinery. Evidence suggests however that the high density of a particular conserved nucleotide sequence motif among tetrapod switch regions (DGYW and preferentially AGCT) may be more important for targeting than the formation of R-loops – a switch region from the amphibian *Xenopus laevis* is A-T rich and not prone to form these structures, yet still efficiently targets CSR when tested in place of an endogenous S region in transgenic mice (Zarrin et al. 2004). While tetrapod *IgH* loci encode several C elements located

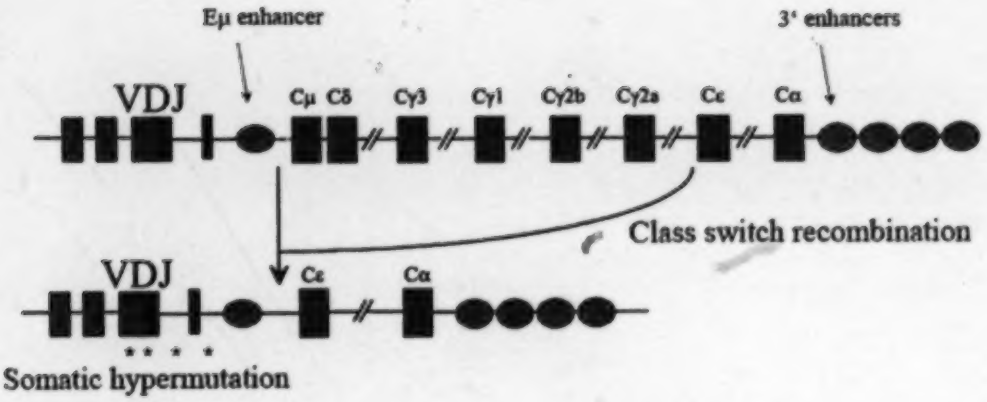
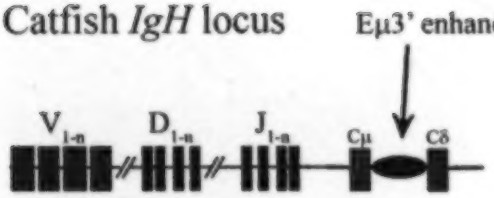


Figure 4 – Schematic (not to scale) of the murine *IgH* locus before and after class switch recombination (CSR). CSR and somatic hypermutation (SH) occur following B-cell activation. During class switch recombination, a portion of the germline *IgH* locus is excised by recombination between switch regions, one of which is located upstream of each C element. This repositions a particular C element in proximity to the rearranged VDJ exon such that it will encode the constant region of the expressed heavy chain peptide. SH involves point mutations introduced to the rearranged VDJ exon which can result in changes in the amino acid sequence of the variable region it encodes. These mutations occur at particular sequence hotspots. Transcriptional enhancers (circles) are required for recombination and mutation processes occurring at the *IgH* locus and transcription of the immunoglobulin gene. The main transcriptional enhancer, Eμ, is situated in the J_H-Cμ intron – a position that is proximal to a promoter (not shown) upstream of the rearranged VDJ exon and which is not susceptible to excision during VDJ recombination or CSR.

downstream of Cμ, the number and identity of these exon clusters vary among the vertebrate radiations (Fig. 3). For example, in *Xenopus laevis*, a member of the earliest diverged vertebrate lineage to undergo CSR (reviewed in Du Pasquier 2001), three C elements leading to heavy chains of the IgM, IgX or IgY isotype are present in the germline locus. The mouse *IgH* locus encodes 8, which give rise to the constant regions of the heavy chains of IgM, IgD, IgGγ3, IgGγ1, IgGγ2b, IgGγ2a, IgE, and IgA. The Cμ and Cδ elements, which give rise to IgM and IgD heavy chains respectively, are situated relatively close together and are not selected between by CSR but rather a mechanism of alternate splicing of a long pre-mRNA transcript. There has not yet been a described function for IgD antibodies, and it is interesting to note that this isotype is absent from several vertebrate lineages. Until recently, only two immunoglobulin isotypes had been described in teleost fish – IgM and IgD (Warr 1995; Hordvik et al. 1999; Srisapoome et al. 2004). There is no evidence that fish undergo CSR. One of the best studied teleost fish *IgH* loci is that of the channel catfish, *Ictalurus punctatus*. The heavy chain constant regions of the teleost IgM and IgD isotypes are encoded by two C elements (Cμ and Cδ) in their *IgH* gene (Fig. 5). These C elements are situated close together in a manner similar to μ and δ in the human and mouse *IgH* loci. Alternate processing of a long transcript including the rearranged VDJ exon, Cμ, and Cδ elements allows the co-expression of μ and δ heavy chains. In the catfish, however, the Cδ element located just 3' of Cμ only encodes the membrane form of the heavy chain of IgD. The secreted form appears to be encoded by a separate Cδ gene located well upstream of the V_H elements in the locus (Bengten et al. 2002). The region of the locus containing the Cδ element encoding the secretory form of the IgD heavy chain appears to have arisen through

a. Catfish *IgH* locus



b. Zebrafish *IgH* locus

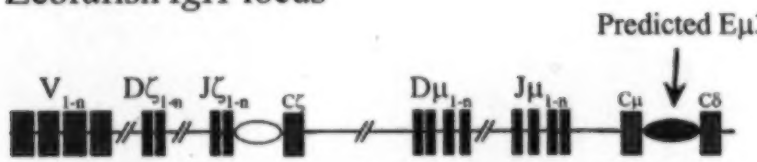


Figure 5 – Schematic (not to scale) representations of the a) channel catfish (*Ictalurus punctatus*) and b) zebrafish (*Danio rerio*) *IgH* loci. Teleost fish such as the catfish are known to express antibodies of the IgM and IgD isotypes, which are encoded by two C-elements, Cμ and Cδ in their *IgH* gene. Like Cμ and Cδ of mammals, co-expression of IgM or IgD heavy chains is achieved by alternate processing of a long mRNA transcript. Recently, analysis of the fully sequenced zebrafish *IgH* locus led to the identification of a new immunoglobulin C-element called Cζ, which encodes the constant region of the heavy chains of IgZ antibodies (Danilova et al. 2005). Transcripts of IgZ heavy chains always contain sequence from proximal D and J elements (Dζ, Jζ), as do IgM/IgD heavy chain transcripts (Dμ, Jμ). There is some evidence that a Cζ exon cluster or remnant thereof is conserved among several teleost species including trout, atlantic salmon, carp and pufferfish, but the location in the catfish *IgH* locus where it would be expected to reside has not yet been analyzed. The locations of transcriptional enhancers are denoted

by filled circles. The catfish *IgH* locus contains a transcriptional enhancer between the Cμ and Cδ elements, downstream of the location of the Eμ enhancer in the higher vertebrates (see Fig. 4). A predicted Eμ3' enhancer in the zebrafish locus is indicated. A possible J_H-Cζ enhancer in the zebrafish locus is represented by an unfilled circle.

duplications of the downstream Cμ and Cδ elements. Recent complete sequencing and analysis of the germline zebrafish (*Danio rerio*) *IgH* locus gave rise to the discovery of a new immunoglobulin isotype called IgZ (Danilova et al. 2005), the C-element for which (Cζ) is located upstream of the Cμ and Cδ elements but downstream of the V elements (Fig. 5b). The function of IgZ is not known, however it is expressed earlier in development than IgM and it appears that in IgM expressing B-cells it is spliced from the locus via VDJ rearrangement to D and J elements upstream of the Cμ constant region and not CSR (Danilova et al. 2005). IgZ-like transcripts from Rainbow trout (*Oncorhynchus mykiss*), Atlantic salmon (*Salmo salar*), and Carp (*Cyprinus carpio*) have also been identified, and a region upstream of Cμ with homology to the zebrafish Cζ element was identified in Japanese pufferfish (*Fugu rubripes*) genomic sequence (Danilova et al. 2005). The equivalent region of the channel catfish *IgH* locus has not yet been described.

During somatic hypermutation at germinal centers in mice and humans, point mutations are introduced to the rearranged VDJ exon (Fig. 4), which tend to accumulate at particular sequence hotspots that are the same as those implicated in targeting CSR (reviewed in Betz et al. 1993; Rogozin and Diaz 2004). The process is coupled with selection for those B-cells which accumulate beneficial mutations leading to production of Ig with higher affinity for its cognate ligand. B-cells which lose the ability to bind their specific antigen via mutation are likely programmed to die by apoptosis. Evidence of SH has been found in cartilaginous fish (sharks), and data seem to indicate that the mutational machinery involved in the SH process was present in the ancestor of all jawed vertebrates (reviewed in Flajnik 2002; Litman et al. 2004). The necessity of germinal centers in mediating affinity maturation is an unresolved issue. Germinal centers have

only been shown to occur in warm-blooded vertebrates (birds and mammals). The extent to which affinity maturation occurs in the ectotherms is very limited and thus is consistent with a contribution of these structures in facilitating SH. The birds, however, which do form germinal centers, appear to show a poor affinity maturation response (reviewed in Du Pasquier et al. 1998).

Activation induced cytidine deaminase (AID) is an enzyme required for CSR, SH and GC (Muramatsu et al. 2000; Revy et al. 2000; Arakawa et al. 2002). Its precise mechanism of function in all of these processes is not yet clear, however it functions to deaminate cytidine residues in single stranded DNA with a preference for the sequence RGYW. Evidence suggests that CSR, SH, and GC are effected by differential processing of AID-generated lesions (Di Noia and Neuberger 2002; Rada et al. 2002; Imai et al. 2003). The targeting of AID to particular sequence hotspots during SH and CSR has been shown to depend on association with a ssDNA binding protein called replication protein A (Chaudhuri et al. 2004; Zarrin et al. 2004) which is involved in DNA repair and replication (reviewed in Binz et al. 2004). AID has been cloned from the channel catfish, *Ictalurus punctatus* (Saunders and Magor 2004) and zebrafish (Zhao et al. 2005) and is clearly present in other teleosts. While the basic machinery for SH may be present, teleost fish show a poor affinity maturation response compared to mice or humans (Cain et al. 2002) although there is some evidence of somatic mutation and selection (Kaattari et al. 2002), and mutations appear to accumulate in the VDJ exon of stimulated catfish cells (Saunders and Magor, unpublished observations). Differences in the identity and layout of functional elements among teleost and higher vertebrate loci are predicted to be

key to understanding how the *IgH* locus evolved and the variations in the extent to which affinity maturation occurs among the lineages.

Promoters, enhancers, and transcription factors control gene expression

The control of gene expression by *cis* regulatory elements is the subject of much research and many of the methods by which it occurs have been well described (reviewed in Carey and Smale 2000). *Cis*-acting elements are typically clusters of transcription factor binding sites in DNA. Bound transcription factors facilitate loading of RNA polymerase and the subsequent transcription of a gene. There are several types of *cis* elements that can regulate gene transcription, including promoters, enhancers, and insulators. Promoters are *cis* acting elements that act as sites of assembly for the RNA polymerase II holoenzyme (reviewed in Smale and Kadonaga 2003). They are located within a few hundred base pairs of the transcription start site and define from where and in which direction transcription proceeds. Many promoters share a set of common core elements, called the core promoter. The positioning of these core elements determines the site at which transcription begins. As one example, a typical core promoter contains a TATA box generally located 25 - 30 bp upstream of the transcription start site. TFIID, a general transcription factor involved in RNA polymerase II transcription initiation, binds to the TATA box and facilitates the binding of the other general transcription factors and the RNA polymerase II holoenzyme. A promoter may or may not be sufficient to regulate the expression pattern of a gene itself. Some genes rely on additional *cis* acting elements in addition to promoters to aid in regulating the expression of a gene.

Enhancers are *cis* acting elements that can serve to regulate the rate of loading of a functional transcription complex at the promoter, and thus the rate of transcription (reviewed in Blackwood and Kadonaga 1998). Enhancers are typically located distal to a promoter with which they interact. While they are sometimes referred to as being distance (relative to the promoter) independent, they tend to be in the vicinity of the promoter, either upstream of it, or in the first intron of the gene. Unlike promoters, enhancers do not depend on a particular orientation in order to be functional. Enhancers are thought to be brought into proximity with promoters through the looping out of intervening DNA such that transcription factors bound at the enhancer can interact with factors bound at the promoter. These interactions can result in stabilization and formation of an active transcription complex.

Transcription factors, or activators, have a multi-component protein structure which in general consists of a DNA binding domain and an activation domain. Many also contain domains involved in mediating protein-protein interactions with other transcription factors. Targeting to specific binding sites in the DNA is achieved through the DNA binding domain. Many different types of DNA binding domains, for example basic leucine zipper (bZIP) and basic helix-loop-helix (bHLH) domains, have been described in eukaryotic transcriptional activators (reviewed in Carey and Smale 2000). Variations in key amino acids among these DNA binding domains underlie the ability of particular DNA sequences to be recognized and bound by them. Specific targeting may also depend on interactions with other proteins.

Activation domains interact with the transcriptional machinery in order to recruit it to or stabilize its interaction at the promoter (reviewed in Carey and Smale 2000). This

may occur directly or through interactions with coactivators which bind the polymerase complex and the transcription factor. Some transcriptional activators bind DNA as multimeric complexes, most typically as homodimers or heterodimers with members of the same transcription factor family. There may also be cooperative or synergistic binding of adjacent transcription factors to transcription factor binding sites in the DNA.

Gene transcriptional regulation via promoters and enhancers is achieved in several ways (reviewed in Carey and Smale 2000). One way to regulate gene expression is to control whether a particular transcription factor involved in mediating transcription of a specific gene is actually expressed in the cell. In situations where cooperative or synergistic binding between factors is required for transcriptional activation, the presence or absence of a single factor, even when other required factors are present, can determine whether a gene is expressed or not. Other methods for gene expression regulation can involve regulating the activity of transcriptional activators themselves. Post-translational modifications to transcriptional activators, for example proteolytic cleavage, phosphorylation or acetylation, can influence their activity. Multimeric transcription factors with functional domains organized on different subunits can often be negatively regulated by alternate inclusion of subunits lacking a DNA binding domain or activation domain. For example in mammals the E-proteins, which are bHLH transcription factors, are negatively regulated by formation of heterodimers with the Id proteins. Id proteins lack DNA-binding domains and function as dominant negative inhibitors of E-protein function - heterodimers containing Id cannot bind DNA (reviewed in Engel and Murre 2001). The expression of repressors which can compete with an activator for its binding site on the DNA also contribute to the control of gene expression.

The regulation of access for binding factors to the DNA represents another level at which gene expression is regulated. In the nucleus genes located in condensed chromatin are not accessible to the transcriptional machinery and thus cannot be expressed. Chromatin structure can be modified by chromatin remodeling complexes. The two general classes, histone acetyltransferase complexes and ATP-dependent chromatin remodeling enzymes, provide access to the DNA for transcription factors in the nucleus, and ultimately RNA polymerase and the factors required for transcription initiation (reviewed in Carey and Smale 2000).

The function of transcriptional regulatory elements in the IgH locus

Like any gene, the *IgH* gene relies on the presence of appropriate transcriptional regulatory elements in its vicinity within the genome. However, the somatic germline alterations that occur at the *IgH* locus present special considerations for the organization of such elements. A constraint in the evolution of the processes occurring at the *IgH* locus is predicted to have been the requirement for the appropriate positioning of these regulatory elements, including B-cell specific promoters and enhancers, within the locus such that they would not be excised by germline recombination events or mutated. The *IgH* loci of vertebrates that undergo CSR like *Xenopus*, mouse and human have the key feature of an enhancer positioned in the intron between the 3' most J_H element and the first constant region gene cluster, Cμ, upstream of the Cμ switch region (Fig. 4) (Banerji et al. 1983; Gillies et al. 1983). This main transcriptional enhancer, called Eμ, is proximal to promoters located upstream of the V_H elements in the locus, and is required for normal *IgH* expression.

In mice another weaker enhancer called DQ52 is located immediately upstream of the most J-proximal D element in the unrearranged germline locus (Kottmann et al. 1994). This enhancer is excised during VDJ recombination and would thus not be present to aid in transcription of a rearranged *IgH* gene.

Four additional enhancers are also located at the 3' end of the locus. These enhancers, HS3a, HS1-2, HS3b, and HS4, are involved in mediating class switch recombination to particular isotypes. In transgenic mice, deletion of all four of these enhancers had little effect on the expression of Cμ-containing IgH transcripts but significantly decreased the expression of Cγ3, Cγ2b, Cγ2a, and Cε (Dunnick et al. 2005). These 3' enhancers may be too distal from the V_H promoters to associate with them except possibly after a class switch to the Cα constant region (see Fig. 4), which may explain the lack of significant effect of their deletion on Cμ expression. CSR to particular C elements is influenced by extracellular signals such as cytokines, which correspond to the nature of the immune challenge (reviewed in Manis et al. 2002). In general, the outcome of CSR is thought to be controlled by changes in the activity of S region promoter- and 3' enhancer- bound factors in response to a particular cytokine profile.

The J_H-Cμ intron is an enhancer location that while in proximity to V_H promoters, is not susceptible to excision during VDJ recombination or class switching (see Fig. 4). While SH is linked to transcription of the *IgH* gene, it is also generally held that some feature of the Eμ enhancer specifically recruits the AID mutator complex to the locus because other enhancers cannot substitute for Eμ in *in vitro* SH assays (Bachl et al. 1998)

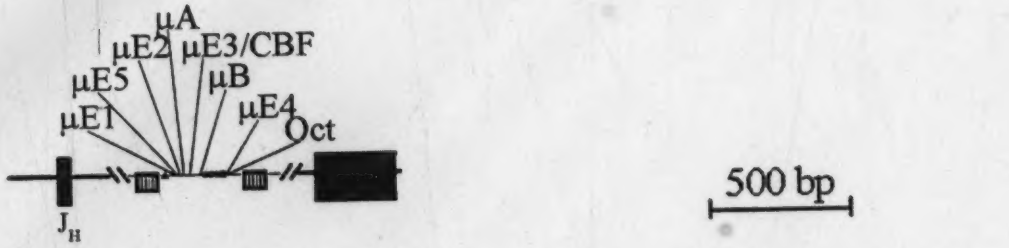
and the rate of SH at the *IgH* locus is at least an order of magnitude greater than at other B-cell expressed genes (Shen et al. 1998).

Basic structure of the Eμ enhancer

B-cell specific Eμ activity depends on a complex and partially understood set of combinatorial interactions between *cis* and *trans* acting elements at the enhancer. These interactions result in the nucleation of an enhanceosome complex that can activate transcription of the *IgH* locus (reviewed in Calame and Sen 2004). There is also evidence that some factors binding at the Eμ enhancer can induce chromatin accessibility. For example in T-cells, ectopic expression of PU.1 or E47 (discussed below) could induce Eμ activity when the *IgH* locus would be expected to be located in silent chromatin (Choi et al. 1996; Nikolajczyk et al. 1999). PU.1 binding has also been shown to alter chromatin accessibility near its binding site in Eμ (Nikolajczyk et al. 1999; Marecki et al. 2004). Interactions of sequence specific factors with DNA in condensed chromatin may underlie the mechanism of targeting chromatin remodeling complexes to particular regions of the *IgH* gene.

The murine and human Eμ enhancers are comprised of multiple binding sites for both ubiquitously expressed and tissue-specific transcription factors, dispersed across a region of approximately 200 bp (Fig. 6) (reviewed in Ernst and Smale 1995; Calame and Sen 2004). Many of these sites bind activators or repressors that contribute to enhancer function – cooperation or synergy between binding factors allows fine regulation. Several sites have also been shown to be redundant in function, although a core structure

Mouse Eμ enhancer



Catfish Eμ3' enhancer

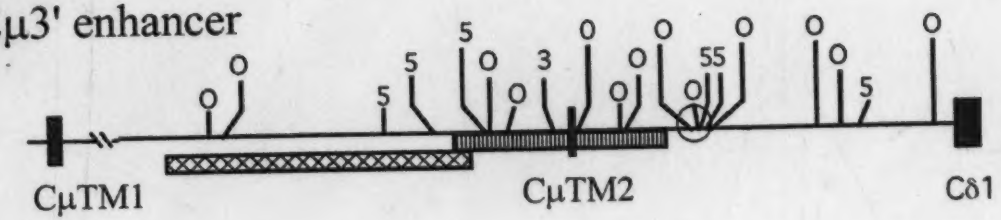


Figure 6 – Diagrams (approximately to scale) of the mouse Eμ and channel catfish Eμ3' enhancers. Enhancer motifs are indicated. Abbreviations: octamer (O), E-boxes μE1-μE5 (1 to 5), core binding factor (CBF). On the map of Eμ, additional CBF binding sites are indicated by small black circles, a binding site for an interferon regulatory factor, IRF-1, is indicated by a small black square, and an E site which binds C/EBP family members is indicated by a small black triangle. Exons are denoted by dark-grey shaded boxes. Predicted matrix attachment regions (MARs) are indicated with vertical striped boxes and the location Tc1-like transposon remnant in the catfish enhancer is represented by a crosshatched box. The core of the catfish Eμ3' enhancer which consists of a mammalian consensus μE5 site (CAGGTG) and two proximal non-canonical octamer motifs (Magor et. al. 1994; Magor et. al. 1997; Cioffi et. al. 2001) is circled.

has been defined which represents a minimal cell type specific enhancer functional *in vitro*.

In the mouse Eμ, five E-box motifs (designated μE1 through μE5) bind ubiquitously expressed but distinct class I basic helix-loop-helix (bHLH) transcription factors, called E-proteins. E-box motifs all share a common core sequence (CANNTG). There are three class I bHLH factor genes in mammals: *E2A, TF12/HEB, and E2-2* (reviewed in Kee et al. 2000; Engel and Murre 2001). *E2A* gives rise to two proteins, E12 and E47, by alternate processing of a primary transcript. The products of these genes generally bind DNA as heterodimers or homodimers and can interact with other classes of bHLH factors present in the nucleus. For example, interaction of ubiquitous class I bHLH factors with tissue restricted Class II bHLH factors can impart tissue-specific gene regulation (Engel and Murre 2001). Class V bHLH factors include the Id factors which lack a DNA binding domain and negatively regulate the activity of class I bHLH factors by partnering with them (Engel and Murre 2001; Quong et al. 2002). The outcome of the various associations is positive or negative regulation at the enhancer based on the ratios of these transcriptional modulators (and their isoforms) expressed by the B-cell. A μA motif is bound by members of the Ets family of transcription factors, such as Ets-1, which are widely expressed and usually require association with other factors for activity (Calame and Sen 2004). A μB motif is bound by PU.1 in B-cells, which is a member of the Ets family expressed in cells of the hematopoetic lineage (Galson et al. 1993; Chen et al. 1995) and is required for B-cell and macrophage development (McKercher et al. 1996). Nelsen *et al* defined a minimal enhancer comprised of the μA, μE3, and μB motifs which spanned 70 bp and could impart B-cell

specific transcriptional enhancer activity, albeit weakly unless tested as an artificially constructed dimer (Nelsen et al. 1993). Mutation of any of the 3 sites abrogated enhancer function indicating a lack of redundancy within this construct and cooperative or synergistic binding between the binding factors. Additionally, expression of PU.1 and Ets-1 could activate transcription from a minimal enhancer construct in nonlymphoid cells. Studies of the interactions between factors binding at the minimal enhancer have indicated that the spacing between the factors can be critical to its function, likely due to a DNA bend induced by PU.1 binding (Nikolajczyk et al. 1996). Disruption of the μE3 site initially established that an interacting factor was critically involved in mediating enhancer activity (Nelsen et al. 1993), though these experiments also resulted in the mutation of an overlapping CBF binding site. The μE3 element binds bHLH factors which include TFE3, TFEB and USF (reviewed in Ernst and Smale 1995). These factors have the feature of a leucine zipper motif in addition to the bHLH motif which can also mediate protein-protein interactions. These so-called bHLH-ZIP proteins may thus be able to form higher order multimers due to the presence of multiple interaction domains (Calame and Sen 2004). Recombinant TFE3 was shown to interact at the μE3 site and positively affect transcription from artificial enhancer constructs when it was overexpressed (Beckmann et al. 1990). Erman and colleagues (Erman et al. 1998) noticed the absence of the μE3 site and TFE3 binding to the human Eμ enhancer but the presence of the overlapping core binding factor (CBF) binding site within Eμ of both mouse and man, although the identity and spacing of many motifs within this element are otherwise very well conserved (Koop et al. 1996). A minimal enhancer derived from the human Eμ containing the μA, μB motifs and CBF binding site was also able to drive

transcription in a B-cell specific manner. Mutagenesis experiments determined that CBF interaction at this site occurred in Eμ of both species and it was at least able to substitute for the function previously attributed to TFE3/μE3 (Erman et al. 1998). In general TFE3 and other bHLH factors binding to the μE3 site may functionally substitute to various extents or serve a negative regulatory role, perhaps in part by competition with CBF for enhancer occupancy.

The minimal tripartite enhancer core capable of B-cell specific function (μA, μE3, μB) requires dimerization in order to drive transcription to any appreciable extent. This implicates the involvement of other elements located within the J_H-Cμ intron in endogenous transcriptional activation. Accordingly, the multiple E-boxes present within the human and mouse Eμ outside of the minimal enhancer core have been shown to participate with the core in driving transcription when assayed *in vitro*. Inclusion of a μE2 site in the mouse minimal enhancer boosted transcription of a reporter gene, yet could not functionally compensate for disruption of the μE3 site (Dang et al. 1998). The four part enhancer still required dimerization for strong activity, but inclusion of a μE5 site in the construct eliminated this requirement. Thus the minimal functional Eμ enhancer construct that is fully active as a monomer in mouse and man is composed of five required elements: μE5, μE2, μA, μB, and μE3/CBF (see Fig. 6). Targeted disruption of any of the sites within this core results in marked loss of activity (Dang et al. 1998), and the more than additive effect of inclusion of these elements implies functional interactions between the factors that bind them.

Elements flanking the minimal Eμ enhancer region that contribute to transcriptional activation

The small region within the endogenous J_H-Cμ intron encompassing the minimal monomeric enhancer core is surrounded by a number of other transcription factor binding sites that have been shown to provide for functional redundancy and subtle regulation (reviewed in Calame and Sen 2004). Several of these elements and the contribution of the factors that bind them to enhancer activity have been studied.

Three CBF binding sites are present downstream of the minimal core region (Fig. 6), and evidence suggests that the presence of these elements can contribute to transcriptional activation (Dang et al. 1998). This is consistent with the previously described binding of CBF with its motif in the core enhancer (Erman et al. 1998).

A site present 3' of the core capable of binding members of the interferon regulatory factor (IRF) family was found to compensate for mutation of the μE3 box in a mouse enhancer construct (Dang et al. 1998). These proteins are transcription factors which control the expression of many interferon regulated genes (reviewed in Mamane et al. 1999). One member of the IRF family, IRF-1, was shown to be capable of binding to the site and contributing to transactivation when coexpressed with E47 and PU.1 (which bind μE5 and μB, respectively). It is possible that other IRF family members may also be able to interact with the element, which is conserved among the human and mouse enhancers.

An octamer motif is also located near the core region. The involvement of octamer motifs and their interacting factors in mediating immunoglobulin V_H promoter activity has been well described (Matthias 1998). Oct1 and Oct2, members of the POU

family of transcriptional regulators, are present in B-cells where they interact with octamer sites in V_H promoters and activate transcription in cooperation with other factors including the B-cell restricted coactivator OBF-1 (Gstaiger et al. 1995; Luo and Roeder 1995; Strubin et al. 1995). Oct1 is ubiquitously expressed, while Oct2 expression is mainly restricted to lymphoid tissues. Although the octamer site within Eμ is apparently not required for enhancer function *in vivo* (Jenuwein and Grosschedl 1991), which is not surprising given the functional redundancy of Eμ so far described, there is some evidence that it may influence activation in pre-activated versus resting B-cells (Yuan et al. 1995). Oct1 and Oct2 have both been shown to interact with TBP, the subunit of the TFIID general transcription factor that makes contact with core promoter TATA boxes, through their POU domains (Zwilling et al. 1994).

An E site which is unrelated to the μE1-5 boxes previously described is located just 5' of the minimal enhancer core region (Fig. 6). In B-cells this motif is bound by C/EBPβ, a member of the CCAAT/enhancer binding protein (C/EBP) family of basic leucine zipper (bZip) transcription factors which bind similar DNA sequences (Landschulz et al. 1988). There are several members of this evolutionarily conserved family which are involved in mediating activation at a variety of promoters and enhancers (reviewed in Lekstrom-Himes and Xanthopoulos 1998), including Ig V_H promoters (Cooper et al. 1992) where 3 isoforms of C/EBPβ interact with both an E site and Oct-1/Oct-2 to modulate transcriptional activation (Hatada et al. 2000). The three isoforms are expressed by usage of alternate start codons via a ribosome slipping mechanism. The shortest of these isoforms lacks an activation domain but retains its DNA binding domain and ability to dimerize with the others, thereby functioning as a

negative regulator (Descombes and Schibler 1991). Other family members have similar homo- and heterodimerization abilities. For instance C/EBPγ (also called Ig/EBP) which was originally purified based on association with the E site in the Eμ enhancer (Roman et al. 1990), was shown to heterodimerize with C/EBPβ and negatively regulate its activity (Parkin et al. 2002). Interestingly, C/EBPβ has also been shown to interact with PU.1 *in vitro* (Nagulapalli et al. 1995) which would be consistent with a function in the Eμ enhancer. Other immune-function related genes are also regulated in part by C/EBP proteins. Mice deficient in C/EBPα show a loss of expression of the granulocyte colony stimulating factor (G-CSF) receptor and a lack of neutrophils or eosinophils (Zhang et al. 1997). The interaction of C/EBP proteins with the E-site in the J_H-Cμ intronic enhancer may thus also represent a variably positive or negative contributor to the activity of the endogenous enhancer depending on the ratios of the various family members and isoforms thereof present in the nucleus.

HMGA1 (earlier known as HMGI(Y)) is a transcriptional coactivator that belongs to the HMGA family of proteins (reviewed in Reeves and Beckerbauer 2001). These factors have widespread involvement in transcriptional regulation and share the common feature of A-T hook protein-DNA interaction domains which bind to the minor groove of AT rich tracts of DNA. They also contain multiple protein-protein interaction domains. The HMGA family is evolutionarily well conserved and homologues exist in species from plants to humans. HMGA1 is present in multiple isoforms (HMGA1a, b, c) which are derived by alternate splicing. It has been shown to interact with PU.1 *in vitro* (Nagulapalli et al. 1995) as well as Ets-1 (McCarthy et al. 2003) and facilitate DNA binding by these factors. Chromatin immunoprecipitation studies have also revealed its

association with the Eμ enhancer in B-cells, where it has been shown to aid in transcriptional activation but does not directly bind to Eμ enhancer DNA (McCarthy et al. 2003). This is not the only context in which HMGA1 has been shown to aid in complex formation without DNA binding, however HMGA1 and related proteins have well established DNA binding and conformation altering properties (Grosschedl et al. 1994). At the interferon-β (IFN-β) enhancer, HMGA1 interacts with motifs in enhancer DNA and is essential for assembly of the enhanceosome (Thanos and Maniatis 1995). The conformational changes its binding induces to the DNA and the protein-protein interactions that it forms has been shown to be key to transcriptional complex assembly at this tightly regulated gene (Yie et al. 1999). HMGA1 has been shown to be subject to acetylation, which can regulate its function (Munshi et al. 1998).

As previously discussed, transcriptional regulation in general is highly influenced by the chromatin configuration around a gene. The widespread chromatin remodeling that must occur in order for some Eμ binding proteins to have access to their sites of interaction on the DNA, as well as for the enhancer to be capable of activation in conjunction with a distal regulatory element (the promoter) is in part facilitated by another key functional element conserved among *IgH* enhancers. Flanking the Eμ enhancer are two matrix attachment regions (MARs) - A-T rich DNA sequences that bind proteins which organize DNA on or about the nuclear matrix. Bright (B cell regulator of *IgH* transcription) is a member of a family of evolutionarily conserved A-T rich interacting domain (ARID) containing proteins (reviewed in Wilsker et al. 2002). These proteins, several of which have been shown to be involved in transcriptional and developmental regulation, interact with target DNA sequences through their ARID

domains. Bright binds to the Eμ MARs as a multimer (Herrscher et al. 1995). Immunoprecipitation studies have also established that Bright associates with DNA in complex with Bruton's tyrosine kinase (Btk) in B-cells (Webb et al. 2000), a defect in which is known to cause X-linked immunodeficiency in mice and X-linked agammaglobulinemia in humans (reviewed in Satterthwaite and Witte 2000). A widely expressed factor has also been shown to specifically bind at the Eμ MARs in non B-cells. Cux/CDP (CCAAT displacement protein), interferes with Bright binding to its sites in the MAR and has been suggested to act as a negative regulator of the locus (Wang et al. 1999). By organizing chromatin into domains or loops on the nuclear matrix, MARs have been shown to isolate a gene from the effects of regulatory elements on different parts of a chromosome (Stief et al. 1989). The Eμ enhancer and MARs have also been shown to mediate position independent expression of a GFP transgene in mouse and human B-cell lines (Lutzko et al. 2003). Transgenic assays have demonstrated that the Eμ MAR directs the formation of accessible chromatin, as measured by histone H3 and H4 acetylation, at a distal position from the enhancer (Jenuwein et al. 1997). Other transgenic studies have indicated that deletion of the MAR results in a significantly reduced level of histone H3 and H4 acetylation relative to wildtype near a distal V_H promoter (Fernandez et al. 2001). Thus it appears that the Eμ MARs aid chromatin remodeling factors in generating widespread accessibility at the endogenous *IgH* locus. The regulation of regions of chromatin structure has been implicated in facilitating the gene rearrangement processes occurring at the locus (Chowdhury and Sen 2004).

The immunoglobulin heavy chain enhancer of the Channel Catfish – Eμ3'

Although the teleost fish represent the earliest extant group of organisms to have the translocon (or V_{1-n}-D_{1-n}-J_{1-n}-C_n) arrangement of their immunoglobulin heavy chain genes, there are key differences in the layout of their *IgH* loci compared to those of higher vertebrates (reviewed in Warr 1995). Like higher vertebrates, recombination processes and transcription at the teleost *IgH* locus rely on enhancer elements appropriately positioned within it to drive transcription from V_H promoters. The catfish immunoglobulin heavy chain enhancer was the first teleost *IgH* enhancer to be functionally characterized. It is situated in a region of the *IgH* locus in the intron between the Cμ and Cδ C elements (Magor et al. 1994a), downstream of the J_H-Cμ intronic Eμ enhancer of the higher vertebrates. Although its structure differs significantly from that of Eμ; lacking many of its key elements, it represents an effective B-cell specific enhancer when tested in either mammalian or catfish cell lines.

Structure of the Catfish Eμ3' enhancer

The structure of the catfish Eμ3' enhancer differs markedly from that of Eμ (Fig. 6). The Cμ-Cδ intergenic region of the catfish contains many putative transcription factor binding sites, including several consensus and nonconsensus octamer and E-box motifs dispersed over a much larger region of DNA (~1.8 kbp) than Eμ (Magor et al. 1994a; Magor et al. 1997). Mutational analyses using *in vitro* reporter constructs established that a mammalian consensus μE5 site in the enhancer is a key mediator of its function (Cioffi et al. 2001). It was also noted that the effect of mutation of this element was more drastic when tested in murine B-cells (abrogated enhancer activity) than in

catfish B-cells (~55% reduction in activity). Further mutation of two octamer motifs found proximal to the μE5 site eliminated function of the enhancer in catfish cells, implicating their involvement in enhancer function. The consensus μE5 site and the proximal octamer pair were defined as the core of the enhancer. A predicted μB motif did not appear to contribute to enhancer activity when tested in either catfish or mouse B-cells (Magor et al. 1997). Mutations of octamers and nonconsensus μE5 motifs within the locus outside the μE5/octamer core were also shown to influence the activity of the enhancer to varying degrees, indicating a complex association between octamer and μE5-box binding factors at the catfish Eμ3' enhancer. With the consensus μE5 box left intact, disruption of a combination of three octamers within the enhancer led to different effects on activity when tested in mouse or catfish B-cells. While the mutated enhancer had ~65% activity in catfish B-cells, it maintained less than 30% in a mouse B-cell line (Cioffi et al. 2001).

Although there are clearly interspecies variations in transcription factor repertoires, the ability of an apparently simpler, more diffuse enhancer to effectively drive B-cell specific transcription in catfish and mouse cell lines indicates a level of evolutionary conservation of such factors among the vertebrates. Indeed, homologues of mammalian E-protein families, which bind the E-box sequences, have been cloned from the channel catfish (Hikima et al. 2004; Hikima et al. 2005b) and identified in *Takifugu rubripes* genomic sequence (Hikima et al. 2005a). In addition, all of the described catfish E-proteins have been shown to bind the μE5 site *in vitro* and drive transcription from the catfish Eμ3' enhancer core.

Implications of the location and structure of Catfish Eμ3'

The Eμ3' transcriptional enhancer of the catfish is located between the μ and δ constant region gene clusters, but in higher vertebrates, the μ and δ constant region genes are excised from the germline locus after a class switch to any other isotype. It has been argued that the Eμ enhancer evolved after teleost fish diverged from the ancestral vertebrate lineage (Magor et al. 1999).

Eμ must have arisen in the J_H-Cμ intron prior to the development of CSR as this is the only location in the locus that is proximal to the V_H promoters and yet not susceptible to excision during CSR or VDJ recombination. Catfish AID has been shown to drive CSR of an appropriate artificial template in catfish B-cells (Muramatsu, Magor and colleagues, unpublished observations). Modification of the *IgH* locus in fish to support CSR would therefore likely present a primary impediment to CSR evolution.

Somatic hypermutation of the *IgH* locus in the higher vertebrates, as was discussed, depends on the presence of the Eμ enhancer. Bachl and colleagues demonstrated that by increasing the distance between the promoter and enhancer, the rate of mutation accumulation in an SH assay was greatly decreased (Bachl et al. 1998). In light of this finding, the promoter-distal location of the Eμ3' enhancer is consistent with observations in fish of poor affinity maturation.

Several putative enhancer motifs (Fig. 6) in the catfish Eμ3' are encompassed by a remnant Tc1-like transposon (reviewed in Magor et al. 1999). It has been argued that if the transposon encompassing several enhancer motifs in the catfish Eμ3' enhancer was present prior to the divergence of teleost fish, it may have served as a vehicle for the repositioning of the enhancer (Magor et al. 1999). These predictions are based on the

premise that the common ancestor to fish and tetrapods had an Eμ3'-like enhancer. However, the channel catfish is the only teleost species in which the *IgH* transcriptional enhancer has been characterized.

Hypothesis and objectives

The catfish Eμ3' enhancer is the only teleost fish *IgH* enhancer that has been studied to date. If an Eμ3'-like enhancer organization was present prior to the divergence of the teleost fish and not an anomaly of the catfish, we would expect to find a similar enhancer in other members of this vertebrate group.

Zebrafish (*Danio rerio*) are a popular early-vertebrate model system in biology. They are relatively easy to maintain and breed, have transparent embryos, mature quickly, and are inexpensive. Furthermore, the zebrafish is the subject of genome sequencing efforts that are nearing completion.

The hypothesis to be tested is that zebrafish, like catfish, have an Eμ3'-like enhancer which resides in the Cμ-Cδ intergenic region of the immunoglobulin heavy chain gene and no Eμ-like enhancer. To test this hypothesis, regions of the zebrafish *IgH* locus were cloned into a chloramphenicol acetyl-transferase (CAT) reporter vector containing a fish V_H promoter and transiently transfected in mouse and catfish B-cell and T-cell lines.

The objectives of this research were:

1) To determine whether a B-cell specific and cross-species active Eμ3'-like enhancer exists in the Cμ-Cδ intergenic region of the zebrafish *IgH* locus.
2) To determine if an Eμ-like J_H-Cμ intronic enhancer is a feature of the zebrafish *IgH* locus.
3) To identify regions of minimal and maximal enhancer activity within the enhancer, and to correlate the activity with the locations of predicted enhancer motifs.

4) To determine the contribution of particular motifs to the function of the enhancer.

5) To determine if a remnant transposon is present within the zebrafish Cμ-Cδ intergenic region.

Materials and Methods:

Cloning and sequencing of regions of the Zebrafish IgH locus

A PAC (P1 artificial chromosome) genomic clone known to encompass a portion of the Zebrafish immunoglobulin heavy chain gene was the gift of Chris Amemiya (Virginia Mason Research Center, Seattle, WA). The clone was identified from a P1 library created from MboI partial digests of High MW DNA from the blood cells of approximately 200 wild type Zebrafish ligated to the BamHI sites of pCYPAC6 (GenBank AF133437). Average insert size was 115 kbp. Using genomic DNA probes designed using zebrafish immunoglobulin μ heavy chain transcripts (Genbank AF406819, AF281480), restriction fragments of the PAC containing sequence 5' of Cμ1 and 3' of TM1 were identified by Southern hybridization. Primers to amplify the regions from J_H-Cμ1 and CμTM1-Cδ1 by PCR from the PAC template were designed using sequence derived from shotgun subcloning of these fragments:

TM1+ 5'- TCATTTTCCTGTTCCTCATCAC -3';

Cδ1- 5'- TGGAGCTTTAACAACACATGAC -3';

J_H+ 5'- GAAGCGGCACCAAAGTCTCC -3';

Cμ1- 5'- ACTGAGACAAACCGAAGACTG -3'.

The resulting products were subcloned and sequenced. Large clones were sequenced in their entirety with the aid of the GPS-1 (genome priming system) kit (New England Biolabs Inc., Beverly, MA) according to supplied protocols. Using this method, a "primer island" of known sequence is randomly inserted into the plasmid via a transposition reaction. Sequence data was assembled using the Chromatool (Biotools Inc., Edmonton, AB) software package. Sequence assemblies covered the entire region

of each insert to a depth of at least 3 reads (J_H-Cμ1 intron , Genbank accession no. AY682724; μ-chain transmembrane exon 1 (TM1)-Cδ1 exon region, AY682723).

Reporter plasmids

Constructs to test enhancer function of regions of the Zebrafish *IgH* locus were based upon a modified pCAT vector (Promega) containing the full promoter of a goldfish V_H gene upstream of the chloramphenicol acyltransferase (CAT) gene with a polylinker downstream of the CAT gene as previously described (Magor et al. 1994b), with the exception that our version of the reporter has a reversed polylinker orientation. Fragments of the Zebrafish *IgH* locus were amplified by PCR using linkered primers (Table 1) and subcloned into the corresponding restriction sites in the polylinker of the pFVH-CAT reporter (Fig. 7), with the exception of constructs generated using primers P-101 through P-108 where amplification products were blunt-end ligated into the *Eco*RV site of pFVH-CAT. Sequencing was performed to verify insert identity and orientation, although we noted no difference in activity between constructs containing identical enhancer test fragments in either the forward or reversed orientation. Two of the test enhancers were constructed piecemeal. Construct no. 9 was produced by subcloning the PCR product of P-210/P-255 amplification (Table 1) into the *Sac*I site of construct no. 5 and selecting for orientation. This construct contains a *Sac*I site and a 20 bp duplication of bases 1647-1671 of Genbank accession AY682723 at the site of insertion as a result. The "no-core" construct was produced by subcloning the PCR product of P-210/P-246 amplification (Table 1) into *Sac*I/*Sph*I-digested construct no. 6. This construct therefore contains the entire region defined by primers P-210/P-213 (Table 1), excluding bases

1759-1963 of Genbank accession AY682723. A construct to test whether including the zebrafish J_H-Cμ1 intron with an active enhancer fragment from the Cμ-Cδ intergenic region would boost its activity was constructed by digesting the zebrafish J_H-Cμ1 reporter construct with *Sma*I and blunt-end ligating in the PCR amplified fragment defined by primers P-102/P-106 (Table 1). This resulted in insertion of the P-102/106 fragment directly 3' of the J_H-Cμ1 fragment in this reporter. Because the V_H promoter is inactive in T-cells, the modified pCAT vector previously described could not be used to verify cell type specificity. The most transcriptionally active zebrafish *IgH* enhancer fragment (construct no. 3) was blunt end cloned into the pCAT promoter (Promega) vector at the *Sal*I site. This reporter plasmid contains the constitutively active SV40 promoter upstream of the CAT gene. The ability of the enhancer fragment to drive transcription in the context of the SV40 promoter was verified in catfish 1B10 B-cells prior to testing activity in T-cells. The pGL3-Control vector (Promega) is a constitutively active luciferase reporter that was co-transfected along with experimental reporter constructs to allow correction for transfection efficiency. The pBluescript KS+ cloning vector (Stratagene, La Jolla, California, USA) was used as inert carrier DNA to standardize the amount of DNA in each transfection.

Table 1
Primers used for generating Zebrafish *IgH* enhancer constructs.

	5' to 3' sequence [a]	Location	constructs
P-101 (+)	Sac I-ATTTTCCTGTTCCTCATCAC	3-22 [b]	1 of Eμ3'
P-210 (+)	Sac I-GTGCCTCGCTTACACATTGC	404-423 [b]	7, 9, 'No Core'
P-102 (+)	Sac I-GATTATTATTTTCTAGTC	502-519 [b]	2, 8 of Eμ3'
P-103 (+)	Sac I-TGCCTCTCAGATGAACG	1376-1392 [b]	3 of Eμ3'
P-104 (-)	Xho I-CCGTTCATCTGAGAGG	1393-1378 [b]	1 of Eμ3'
P-211 (+)	Sac I-CTGAACACAGCATTGTCTAGC	1723-1743 [b]	4, 6 of Eμ3'
P-246 (-)	Sph I-TTACAAATATGCAGAGCTAGAC	1758-1737	Eμ3' no core
P-249 (+)	Nhe I-ATGTGCTCAGCCTTGTTTCAGA	1851-1872 [b]	Δ5' E-box, Δ5+3' E-box
P-250 (-)	Nhe I-CACAGGAAAAATATCTTATAATTTCA	1863-1845 [b]	Δ5' E-box, Δ5+3' E-box
P-252 (-)	Nhe I-CTGCATCTCTGAAACAAGGCTG	1902-1881 [b]	Δ3' E-box, Δ5+3' E-box
P-251 (+)	Nhe I-GTGCTTTTCGGTTCTGCTTTTG	1903-1924 [b]	Δ3' E-box, Δ5+3' E-box
P-214 (+)	Sac I-GCATGCTTCTGGTTTCCTC	1964-1982 [b]	5 of Eμ3'
P-212 (-)	Xho I-GTGGAAGAGGAAACCAGAAG	1988-1969 [b]	4, 7, 8 of Eμ3'
P-255 (-)	Sac I-GTGGAAGAGGAAACCAGAAG	1988-1969 [b]	9 of Eμ3'
P-105 (-)	Xho I-AGCCATTGAAAGAAGCATTGC	2096-2076 [b]	2 of Eμ3'
P-213 (-)	Xho I-CTTTCAAGTGGCACTCCAAC	3210-3191 [b]	5, 6 of Eμ3'
P-106 (-)	Xho I-ACCGCTGACAGGCTGACC	3269-3252 [b]	3 of Eμ3'
P-107 (-)	Xho I-GAAGCGGCACCAAAGTCTCC	1-20 [c]	J_H to Cμ1 intron
P-108 (+)	Sac I-ACTGAGACAAACCGAAGACTG	2423-2443 [c]	J_H to Cμ1 intron

[a] linkers are indicated with their restriction site name
[b] Position in Genbank sequence #AY682723
[c] Position in Genbank sequence #AY682724

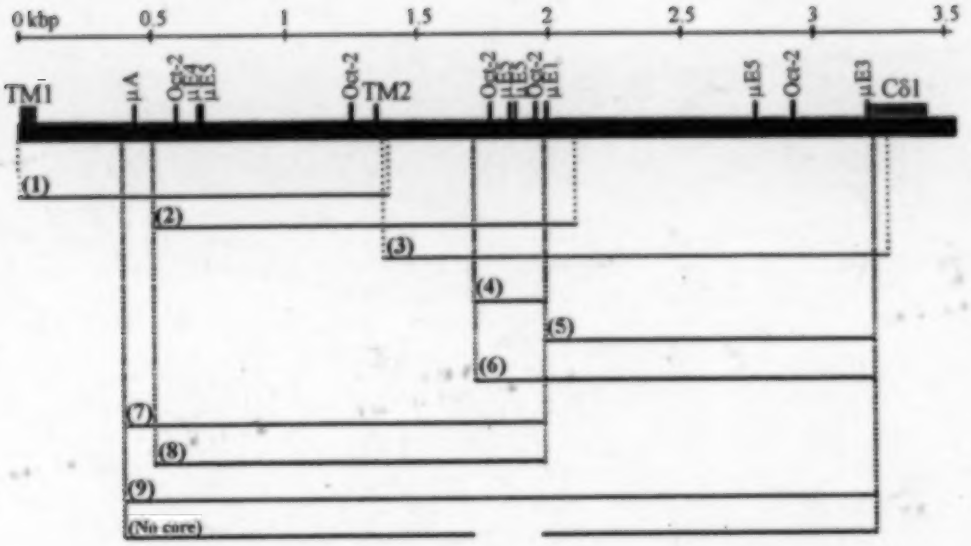


Figure 7 – The relative positions of regions of the zebrafish CμTM1-Cδ intergenic region included in reporter constructs. Indicated on the map are the positions of exons and putative enhancer motifs using designations provided in Table 2.

E-box mutagenesis constructs

Disruption of either the 5' or 3' E-boxes was performed using constructs no. 4 and 7 as a base. The 5' E-box was disrupted by PCR amplification with *Nhe*I-linkered primers P-249/P-250 (Table 1) that abut in the middle of the target motif, using construct no. 4 or 7 as template. Following digestion of the resulting PCR product with *Nhe*I and re-ligation, an *Nhe*I site disrupts the target motif. The 3' E-box was mutagenized in the same manner using *Nhe*I-linkered primers P-251/P-252. Plasmids containing a disruption of both E-boxes (Δ5'+3' E-box) were constructed by PCR amplifying the corresponding Δ5' E-box construct with P-251/P-252, digestion with *Nhe*I, and re-ligation. As a result, the Δ5'+3' E-box constructs differ from those with single E-box disruptions in that they lack 21 bp of sequence between the E-boxes. Plasmids for transfection were purified using either the Qiagen Plasmid Midi or Maxi kits (Qiagen, Valencia, CA).

Growth and transfection of cell lines

The Catfish B cell line 1B10 (Miller et al. 1994) and T cell line 28S.3 (Wilson et al. 1998) were cultured in Catfish medium (45% AIM-V medium, Invitrogen, Burlington, ON; 45% L-15 medium, Invitrogen; and 10% H_2O to adjust for catfish serum tonicity) supplemented with 10% heat-inactivated fetal bovine serum and 1% heat-inactivated carp or goldfish serum. The murine plasmacytoma line J558L was cultured in DMEM (Invitrogen) supplemented with 10% heat-inactivated FBS. The murine T cell line EL4 was cultured in IMDM (Invitrogen) supplemented with 10% heat-inactivated FBS. Each transfection mixture (20 μL) contained 3.4 pmol of experimental reporter, 0.4-2 μg pGL3-Control as a control for transfection efficiency, and pBluescript KS+ as inert

carrier to bring the total DNA mass in each transfection to 20 μg. The amount of pGL3-Control transfected was varied due to differences in the level of luciferase expression among the cell lines for a given mass of plasmid introduced. All transfections were performed using a BTX-ECM 630 electroporator (Biotechnologies and Experimental Research Inc., San Diego, CA, USA) using 2-mm gap cuvettes. Cells were harvested during logarithmic growth, which corresponded to a density of 1-8 x 10^6 cells/mL for 1B10 and 28S.3, and 0.5-2 x 10^6 cells/mL for EL4 and J558L. They were then washed once in their native growth medium without serum, and resuspended in serum-free medium for transfection. 180 μL of cells were combined with the 20 μL transfection mixture just before electroporation. Electroporation parameters were as follows: 1B10: 8 x 10^6 cells/transfection, 210 V, 1100 μF, 50 Ω; 28S.3: 8 x 10^6 cells/transfection, 250 V, 800μF, 50 Ω; J558L: 4 x 10^6 cells/transfection, 133V, 1100 μF, 50 Ω; EL4: 4 x 10^6 cells/transfection, 125V, 2000 μF, 50 Ω. Typical pulse delivery times were ~15 ms. Following electroporation, cells were placed in complete medium supplemented with 25% cell conditioned complete medium in 24-well plates and allowed to recover for 48h.

Reporter assays

Cell extracts were prepared 48h post-transfection using Promega reporter lysis buffer (Promega) according to manufacturers' supplied protocols. CAT activity in lysates was measured using D-threo-[dichloroacetyl-1,2-^{14}C]-chloramphenicol (ICN Biomedicals Inc., Irvine, CA, USA) and n-butyryl coenzyme A (Sigma) according to Promega supplied CAT assay protocols. Luciferase activity was measured using the Promega luciferase assay system (Promega) and a LKB-Wallac 1250 Luminometer.

Values for CAT activity were corrected for transfection efficiency by normalization to luciferase co-reporter expression values. For 1B10 transfections, luciferase values lower than 0.4 after subtracting the mock value were unusually low for this cell type and indicated failed transfections. These were excluded from analyses.

Enhancer activities are presented as a mean percentage of the catfish enhancer (pFVH-CAT-Eμ3') for B-cell transfections ± SEM. Individual values for percent of catfish Eμ3' (ELF11, Magor et. al 1994) activity were calculated by comparison to the activity of a catfish Eμ3' reporter construct transfected in the same set. Reporter constructs were considered to have detectable enhancer activity in 1B10 cells if they showed a mean percent of ELF11 activity higher than 22%, which corresponds to one standard deviation above the mean activity for the enhancerless reporter pFVH-CAT-0. 95% confidence intervals (appendix 1) were used to determine the significance of differences in activity between 1B10 reporter constructs. T-cell transfections and E-box mutagenesis construct transfections are presented as fold induction over enhancerless pSV-CAT or pFVH-CAT reporter construct, respectively, ± SEM. E-box mutagenesis construct transfection fold-induction values were calculated by comparing values for percent of catfish Eμ3' enhancer activity between the experimental reporter and empty vector. Individual fold-induction values for T-cell transfections were calculated by comparison to an enhancerless reporter construct transfected within the same set. A graph of all compiled 1B10 transfection data with 95% confidence intervals (mean ± t*SEM) indicated is included in Appendix 1. Raw data are included in Appendix 3.

DNA sequencing

DNA Sequencing was carried out using the DYEnamic ET terminator cycle sequencing kit (Amersham-Pharmacia Biotech, Uppsala, Sweden) according to standard manufacturer's supplied protocols, with the modification of using half-reactions. Typical sequencing reactions contained ~300-450 ng template, 3 pmol primer and 4 μL sequencing premix in a final volume of 10 μL. Typical thermal cycling parameters were 25 cycles of 95° for 20 seconds, 50° for 15 seconds and 60° for 1 minute. Sequences were run on either an ABI 377 or 373 DNA sequencer. Sequence manipulations were performed using Genetool (Biotools Inc., Edmonton, AB).

In silico analyses of enhancer regions

Analyses to determine the location and identity of putative transcription factor binding motifs were performed with the Transcription Element Search System (TESS, http://www.cbil.upenn.edu/tess) (Petsko 2002). This tool uses the TRANSFAC database of transcription factor binding sites (http://www.biobase.de). Additional *IgH* enhancer motifs defined in the TFD database (Ghosh 1990) were included (see appendix 2). One base mismatch from canonical sequence was allowed.

The zPicture program (http://zpicture.dcode.org, (Ovcharenko et al. 2004)) compares two or more sequences using the BLASTz algorithm (Schwartz et al. 2003). This program was used to identify regions of the *IgH* enhancers conserved among species. "Default" zPicture BLASTz parameters were used, with no repetitive element masking (O = 400 (gap open penalty), E = 30 (gap extension penalty), K = 3000 (threshold for maximal segment pairs), L = 3000 (threshold for gapped alignments), M =

50 (mask threshold for sequence 1)). Using the "fast" (intended for microbial-sized genomes) or "chained" BLASTz options did not change the BLASTz alignment between catfish and zebrafish CμTM1-Cδ regions. Conserved transcription factor binding site searches in aligned sequences were performed using rVista (http://rvista.dcode.org, (Loots and Ovcharenko 2004)), which also references the TRANSFAC database, and by inspection.

Potential matrix attachment regions (MARs) were predicted using the S/MARtest program (http://www.genomatix.de; (Frisch et al. 2002; Liebich et al. 2002)).

Results:

The Zebrafish IgH locus contains a functional Eμ3' enhancer

The J_H-Cμ1 intron and Cμ-Cδ intergenic region of the zebrafish were sequenced in their entirety and analyzed for putative transcription factor binding sites. The sites included in the search were selected based on their established functions in the catfish Eμ3' and mouse Eμ. Analysis revealed a number of putative motifs located within both regions of the locus that may participate in driving transcriptional activation, including several canonical and non-canonical octamer motifs and E-boxes (Tables 2 and 3). Using our *in vitro* assay system, initial functional analysis in catfish 1B10 cells revealed that a portion of the Cμ-Cδ intergenic region downstream of the first transmembrane exon (TM1) was able to drive transcription, but the J_H-Cμ intron showed no higher level of transcriptional activation than empty vector (Fig. 8). Most enhancer activity localized 3' of the second transmembrane exon (TM2). The most active enhancer fragment (construct no. 3, Fig. 8) showed approximately 70% of the activity of the Catfish Eμ3' enhancer in 1B10 cells.

Table 2
Potential transcription factor binding sites in the Zebrafish Eμ3' enhancer

Motif Name[a]	Consensus motif	Motif Start[c]	Actual motif[d]
μA/Ets-1	SMGGAWGY	434	CAGGATGT
Oct-2	ATTTGCAT	591	ATTTGCtT
μE4	ACCACCTG	675	cCCACCTGT
μE5[b]	CAGGTGTT	684	CAGGTtTT
Oct-2	ATGCAAAT	1246	ATGCAAAa
Oct-2	ATGCAAAT	1776	ATGCAAAT
μE5	TGCAGGTGT	1846	TGCAGaTGT
μE5	TGCAGGTGT	1875	TGCAGGTGc
Oct-2	ATTTGCAT	1945	ATTTGCtT
μE1	GCCATCTT	1988	cCCATCTT
μE5	ACACCTGCA	2774	ACACCTcCA
Oct-2	ATTTGCAT	2910	ATTTGCtT
μE3	TGCCACATGA	3198	TGCCACtTGA

[a] As defined in the Transfac database and Ghosh (1990)
[b] There are two sequence variants of this motif – mu-E5 and CμE5
[c] Sequence position in Genbank # AY682723
[d] Bases in lower case are non-canonical

Table 3
Potential transcription factor binding sites in the Zebrafish J_H-Cμ1 intron

Motif Name[a]	Consensus motif	Motif Start[b]	Actual motif[c]
Oct-2	ATgCAAAT	379	ATcCAAAT
Oct-2	ATTTGCAT	663	ATTTGtAT
Oct-2	ATGCAAAT	900	ATaCAAAT
Oct-2	ATTTGCAT	1007	ATTTGtAT
Oct-2	ATTTGCAT	1132	ATTTGCAg
Oct-2	ATTTGCAT	1156	ATTTGtAT
μE5	AACACCTG	1266	AAaACCTG
Oct-2	ATTTGCAT	1737	gTTTGCAT
Oct-2	ATTTGCAT	2100	ATTTcCAT
μE1	AAGATGGC	2157	AAGATGGt
μE4	ACCACCTG	2406	ACCAtCTG

[a] As defined in the Transfac database and Ghosh (1990)
[b] Sequence position in Genbank # AY682724
[c] Bases in lower case are non-canonical

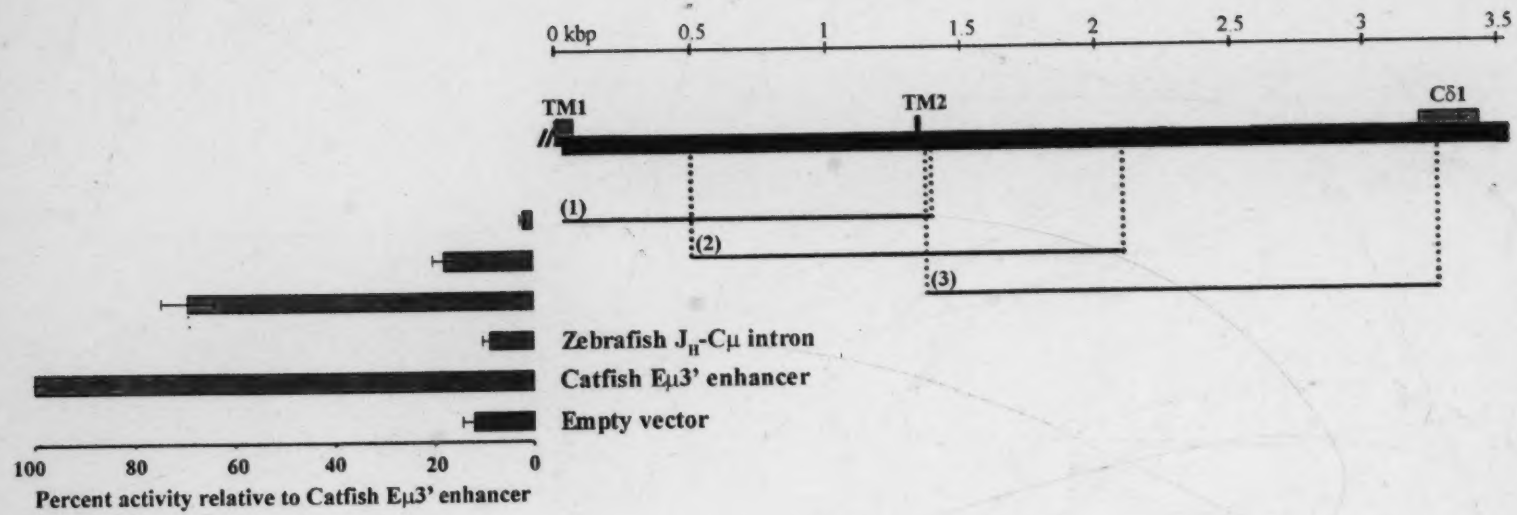

0 kbp
0.5
1
1.5
2
2.5
3
3.5
TM1
TM2
Cδ1
(1)
(2)
(3)
Zebrafish J_H-Cμ intron
Catfish Eμ3' enhancer
Empty vector
100
80
60
40
20
0
Percent activity relative to Catfish Eμ3' enhancer

Figure 8 – Transcriptional enhancer activity from regions of the zebrafish immunoglobulin heavy chain (*IgH*) locus tested in the catfish 1B10 B-cell line. Top: A scaled map of the relative positions of exons (shaded boxes) within the heavy chain Cμ TM1-Cδ intergenic region. Bottom: Regions of the zebrafish *IgH* locus in reporter constructs indicated by lines corresponding to the map (or as indicated in text) were cloned into the pFVH-CAT reporter and transiently transfected into the catfish B-cell line 1B10. CAT reporter activities were measured 48h post-transfection and normalized for transfection efficiency using a co-transfected pGL3 luciferase reporter. Enhancer activities are presented as the mean percentage of the activity (set at 100%) of the catfish Eμ3' enhancer, ± SEM (Magor et al. 1994). The empty vector, pFVH-CAT, has a goldfish V_H promoter but no enhancer fragment inserted in its polylinker. A minimum of eight replicate transfections from three independent experiments are included for each reporter construct.

The Zebrafish and Catfish Cμ-Cδ intergenic regions share a similar enhancer organization

Analysis of the zebrafish and catfish Cμ-Cδ intergenic regions was performed with zPicture (Ovcharenko et al. 2004). Resulting dot-plots indicated a small, ~300 bp region of sequence identity (Figs. 10a, 11), in contrast to the high conservation of the mouse and human J_H-Cμ introns (Fig. 10b). Analysis of the BLASTz aligned sequences for functional motifs described in the catfish enhancer (Fig. 11) revealed that while there was significant sequence conservation, this conservation only appeared to extend to two predicted E-boxes (CANNTG) located 21 bp from each other in each sequence (5' E-box and 3' E-box). In each of the catfish and zebrafish sequences one of the predicted E-boxes is of mammalian consensus μE5 sequence (CAGGTG) and one differs from the canonical μE5 sequence by a single nucleotide (Table 2, Fig. 12a). In the catfish sequence there are two non-canonical octamer motifs located within 50 bp of the consensus μE5-box. The zebrafish sequence has no similarly located octamer motifs but there is one canonical octamer motif located approximately 100 bp upstream of the 3' E-box and a non-canonical octamer 100 bp downstream (Fig. 11). The sequence of the non-canonical octamer motif has been shown to drive transcription in mouse and fish B-cell lines (Magor et al. 1997). In both enhancers, the basic arrangement of a μE5 consensus site in proximity to two octamer motifs appears to be conserved.

The region of conservation between the two Cμ-Cδ intergenic regions encompasses the core of the catfish enhancer (consensus μE5 box and proximal octamer motifs). We therefore wanted to determine if the aligned region also acted as the core of the zebrafish Eμ3'. The predicted zebrafish "core" region (construct no. 4; Fig. 9) had

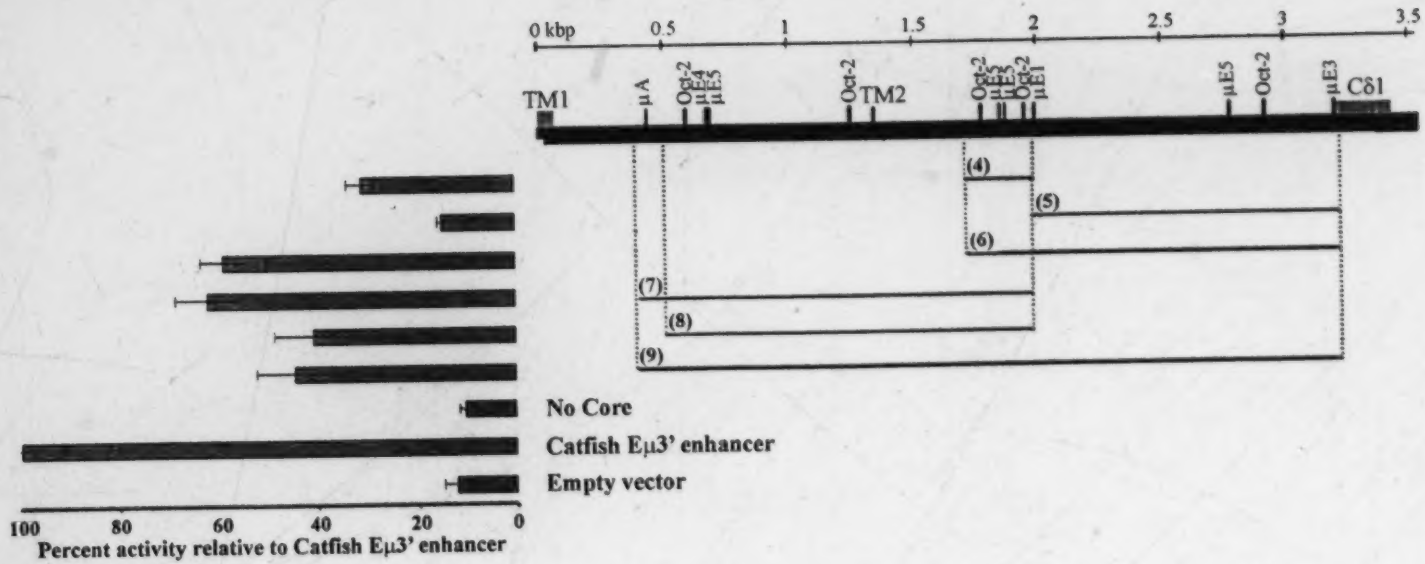

0 kbp
0.5
1
1.5
2
2.5
3
3.5
TM1
μA
Oct-2
μE4
μE5
Oct-2
TM2
Oct-2
μE2
μE3
Oct-2
μE1
μE5
Oct-2
μE3
Cδ1
(4)
(5)
(6)
(7)
(8)
(9)
No Core
Catfish Eμ3' enhancer
Empty vector
100
80
60
40
20
0
Percent activity relative to Catfish Eμ3' enhancer

Figure 9 – Localization of transcriptionally active regions within the zebrafish Cμ-Cδ intergenic region. Top: Scaled schematic of the μ-δ region. The positions of putative enhancer motifs using designations provided in Table 2 are indicated. Bottom: The enhancer activities of fragments of the μ-δ intergenic region (as indicated by lines extending to the map) were tested in catfish 1B10 B-cells as described in Fig. 8. Enhancer activities are presented as the mean percentage of the activity (set at 100%) of the catfish Eμ3' enhancer, ± SEM. Data represent a minimum of eight replicate transfections from three independent experiments.

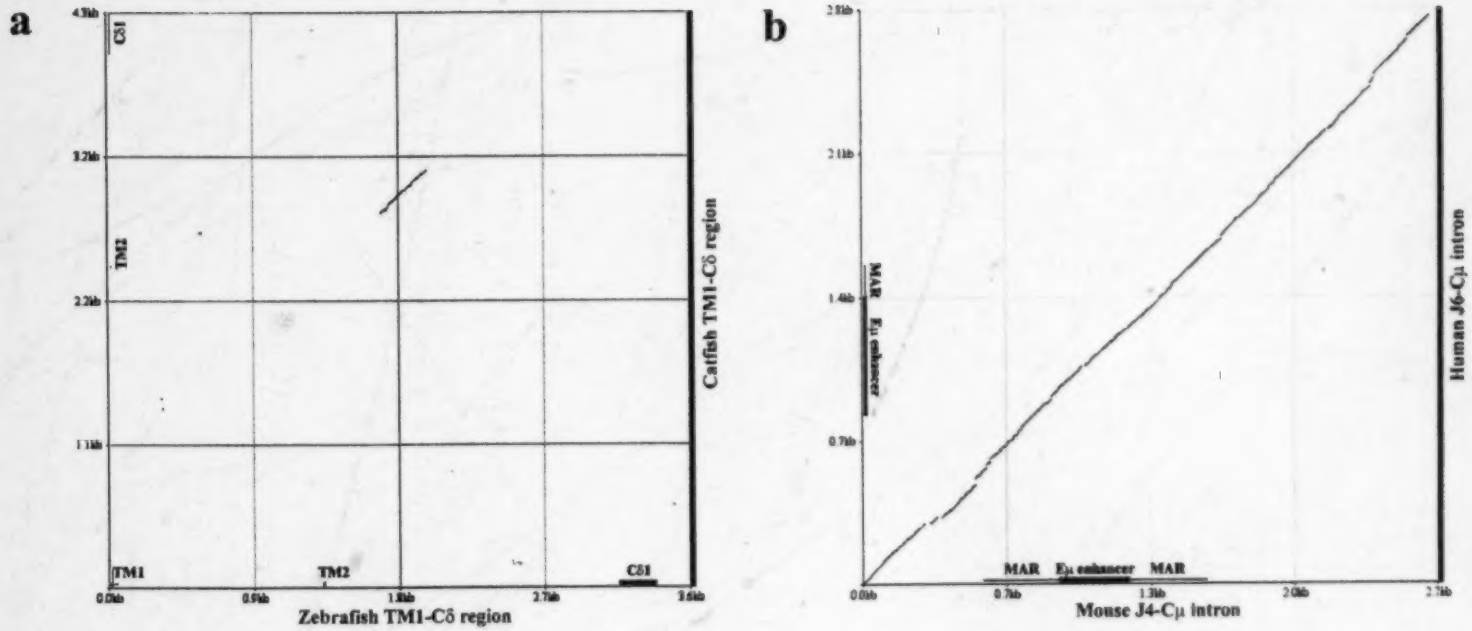

a
Cδ1
TM2
TM1
TM2
Cδ1
Catfish TM1-Cδ region
Zebrafish TM1-Cδ region
b
MAR
MAR
Eμ enhancer
MAR
Human J6-Cμ intron
Mouse J4-Cμ intron

Figure 10 – Dot-matrix comparisons of fish and mammalian *IgH* enhancer regions. zPicture dot-matrix arrays indicate regions of blastz identified sequence homology among the fish (a) and mammalian (b) *IgH* enhancers. The catfish and zebrafish Eμ3' regions were compared from the CμTM1 exons to beyond Cδ1, using "default" zPicture parameters. Exon locations are indicated along the axes. The comparison of the human and murine Eμ regions extends from the most Cμ proximal J element to the beginning of the Cμ switch region in each intron. The position of the Eμ enhancers and flanking matrix attachment region(s) (MARs) are indicated along the axes.

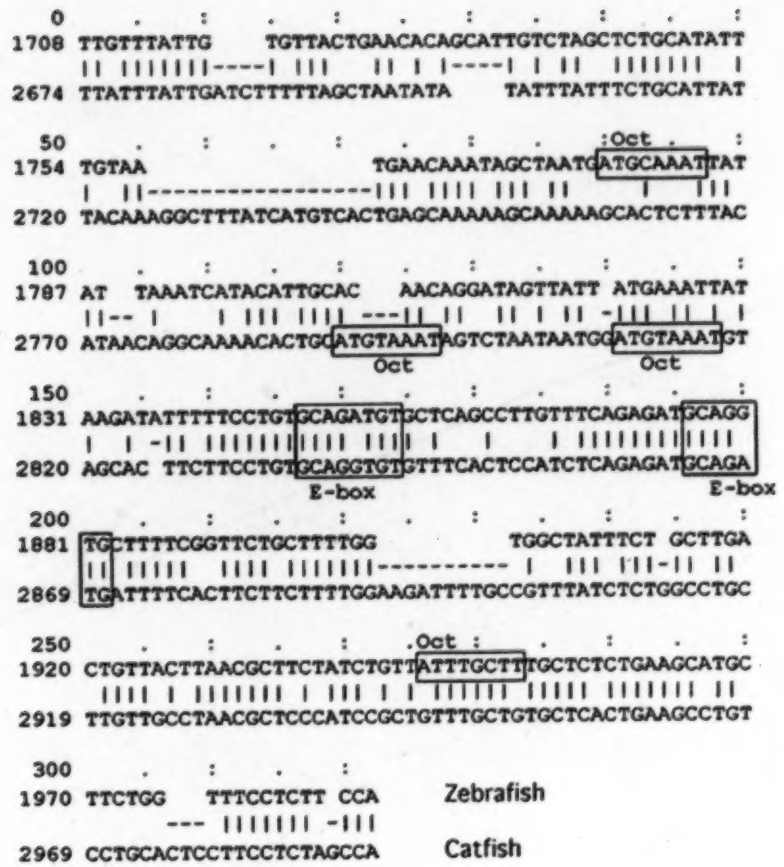


Figure 11 – zPicture generated BLASTz alignment of the zebrafish and catfish Eμ3' regions that have significant sequence similarity. There is 57% nucleotide sequence identity over the conserved region. Base identity is indicated by a vertical bar and gaps are indicated by dashes. Putative enhancer motifs are boxed and labeled. The sequence of the catfish conserved region corresponds to bases 10580-10896 of Genbank accession no. X79482; Zebrafish: 1708-1987 of Genbank accession no. AY682723.

significant activity in our *in vitro* assay, driving transcription to ~31% of the Catfish enhancer. While constructs no. 1 (Fig. 8) and no. 5 (Fig. 9) give little transcriptional activity on their own, constructs containing these core-flanking regions along with the enhancer core (constructs no. 6, 7, 8; Fig. 9) show a substantially higher level of activity than the core alone. In addition, a construct which contained both flanking regions but no core region showed no more activity than empty vector ("no core" construct, Figs. 7 and 9), indicating the requirement for the core region. A similar enhancer organization, with a core and flanking regions that have a more than additive effect on transcriptional activation when included in reporter constructs, was also noted in the catfish Eμ3' enhancer studies.

E-boxes are key elements in the function of the zebrafish Eμ3' enhancer

The finding that a pair of E-box motifs were conserved between the catfish and zebrafish Eμ3' enhancer cores indicated that these E-boxes may be functional elements of the zebrafish enhancer. Furthermore, when predicted Cμ-Cδ intergenic region sequences from Atlantic salmon (Genbank accession nos. Y12391 and AF141605) and Japanese flounder (Genbank accession no. AB158556) were compared to the zebrafish and catfish sequences no significant sequence conservation was identified. However, analysis of putative transcription factor binding motifs in the salmon and flounder sequences revealed the presence of two E-boxes with the same spacing as the E-boxes found in the zebrafish and catfish enhancer cores (Fig. 12a). In order to determine their contribution to enhancer function, mutagenesis was performed to disrupt either the 5' or 3' core E-boxes of the zebrafish Eμ3' enhancer. Mutation of a single E-box (5' or 3') in

a construct containing the core and the 5' flanking region (construct no. 7) resulted in a significant decrease in enhancer activity, while disruption of both E-box motifs completely abrogated enhancer function (Fig. 12b). Similar results were obtained when the same motifs were disrupted in the construct containing only the enhancer core (construct no. 4, Fig. 12b).

a

```
Ip      ---TtCTTCCTGTGCAGGTGTGT--TTCACTCCATCTCAGAGATGCAGATGATTT--
Dr      ATATTTTTCCTGTGCAGATGTGC-TCAGCCTTGTTTCAGAGATGCAGGTGCTTTTCG
Po      ---TCAAGCCTCAACATCTGTGCAACAGACTTCCAGCTCTATTTCCACCTGACTTT-
Ss        TCCTGCTCTAACACCTGCGTT--TCACAACCCTCCAACTCCACATCTGGCCA
```

b

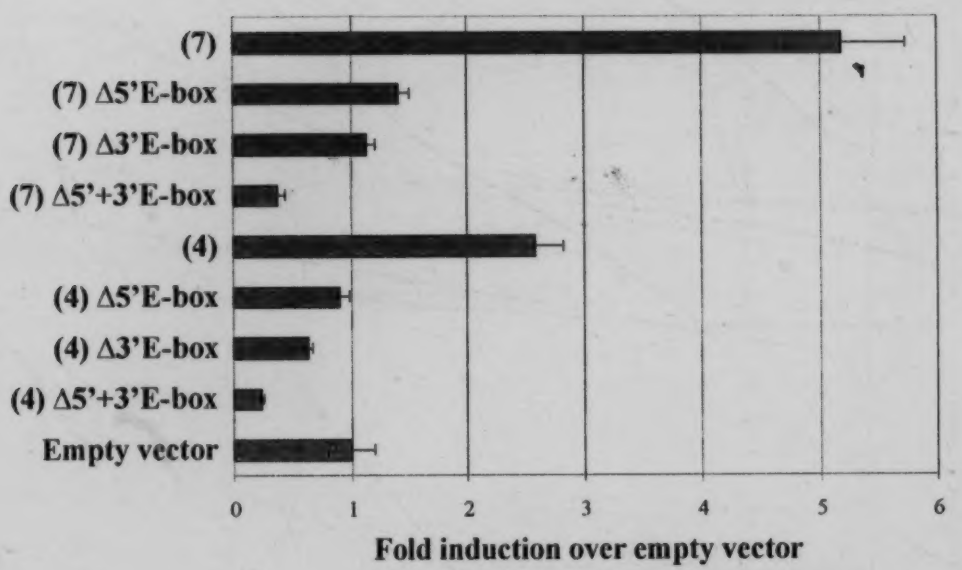


Figure 12 – A pair of conserved E-box motifs are important in mediating transcriptional activity from the zebrafish Eμ3' enhancer. a) A pair of closely spaced E-box motifs (underlined) are found within the catfish (*Ictalurus punctatus, Ip*) and zebrafish (*Danio rerio, Dr*) Eμ3' regions and the predicted Eμ3' regions of Japanese flounder (*Paralichthys olivaceus, Po*, Genbank accession no. AB158556) and Atlantic salmon (*Salmo salar, Ss,* Genbank accession nos. Y12391 and AF141605). b) Either the 5' E-box (Δ5' E-box) or 3' E-box (Δ 3' E-box), when disrupted, cause a significant reduction in enhancer activity driven by an extended enhancer fragment (construct no. 7, see Fig. 9)

or a fragment containing the core alone (construct no. 4, see Fig. 9). Data represents a minimum of six replicate transfections for each construct performed in the course of at least two independent experiments. Results are presented as fold induction over the enhancerless control reporter.

Elements flanking the zebrafish Eμ3' core region contribute to enhancer function

Although regions flanking the enhancer core were shown to participate with the core in mediating transcriptional activation, the functional elements in these flanking regions are unknown. In the 5' core flanking region, there is a predicted Ets-1 binding site (μA). Although mutagenesis of this motif was attempted, the sequence surrounding the site proved unsuitable for design of successful primers. We observed higher enhancer activity in a construct containing the predicted μA site (construct no. 7, Fig. 9) versus a similar construct truncated on the 5' end by 98 bp including the μA site relative to construct no. 7 (construct no. 8, Fig. 9). These data may indicate a function for the μA site, however this would be unexpected as a μA site in the catfish Eμ3' was not shown to contribute to enhancer activity in catfish cells (Magor et al. 1997). A further unexpected observation was made when we cloned the entire TM1-Cδ intron into the pFVH-CAT reporter. This construct (construct no. 9) drove transcription to an equal or lesser extent than constructs containing only the enhancer core and a single core flanking region (either the 5' or 3' flank, see Fig. 9). There is no readily apparent explanation for this finding. It is possible that this construct encompassed unidentified silencing elements, or its piecemeal construction somehow disrupted the activity of functional elements, perhaps by altering the spacing between them. We also tested a reporter construct in which the J_H-Cμ1 intron was situated proximal to an active Eμ3' enhancer fragment (defined by primers P-102/P-106, see Table 1) and noted no increase in transcriptional activity relative to a P-102/P-106 fragment-containing enhancer construct alone (P-102/P-106, 68% of catfish Eμ3 activity; J-Cμ+P-102/106, 67% activity, see appendix 1).

Because mammalian Eμ enhancers have been shown to be associated with matrix attachment regions (MARs), we wanted to determine if the fish Eμ3' enhancers were also associated with these elements. An analysis of the catfish and zebrafish Eμ3' enhancer regions using S/MARtest identified predicted MARs within each of the enhancers. The positions of these MARs relative to the enhancer core differed between the two species (Fig. 13). In the catfish locus, a 940 bp MAR is predicted immediately 5' of the enhancer core. In the zebrafish, two MARs of 350 bp and 540 bp, respectively, were predicted to be located 3' of the enhancer core. Analysis of the J_H-Cμ1 introns of either fish did not identify any predicted MARs.

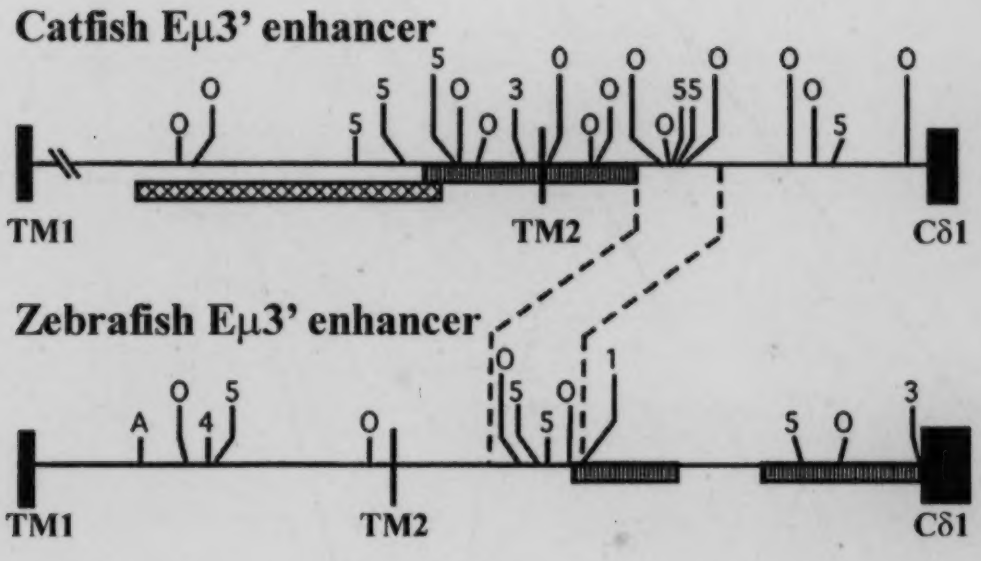


Figure 13 – Scaled maps of the catfish and zebrafish Eμ3' enhancers showing the relative positions of predicted enhancer motifs. The regions of sequence similarity detailed in Figures 10 and 11 are indicated by dashed lines. Exons are indicated by dark boxes. Enhancer motifs are octamer (O), μE1-μE5 (1 to 5), μA (A). Below each map the positions of a Tc1-like transposon remnant in the catfish Eμ3' region (crosshatched box) and predicted matrix attachment regions (MARs) (striped boxes) are indicated.

The zebrafish Eμ3' enhancer can drive transcription in murine B-cells

In order to determine whether the zebrafish Eμ3' enhancer could drive transcription in mammalian B-cell lines, we transfected mouse J558L plasmacytoma cells with enhancer reporter constructs (Fig. 14). In murine B-cells the zebrafish Eμ3' enhancer was functional and appeared to rely on similar interactions of a core and flanking regions to drive transcription. The effect of inclusion of flanking regions with the enhancer core was less pronounced in murine B-cells compared to catfish B-cells, indicating differential activities of the enhancer in catfish and murine B-cells.

The zebrafish Eμ3' is B-cell specific

Because an *IgH* enhancer should be B-cell specific, we wanted to determine if the Zebrafish Eμ3' enhancer could drive transcription in T-cell lines. Because the fish V_H promoter used in our main reporter constructs is inactive in T-cells, we tested the ability of Eμ3' to drive transcription in the context of an SV40 promoter which is active in B and T-cells. We found that, as expected, the zebrafish enhancer did not drive transcription in the catfish T-cell line 28S.3 nor the murine T-cell line EL4 (Fig. 15). The enhancer fragment was able to drive transcription in the context of the SV40 promoter when tested in B-cell lines (1B10: 2-fold induction over enhancerless reporter, J558L: 3.8-fold induction over enhancerless reporter).

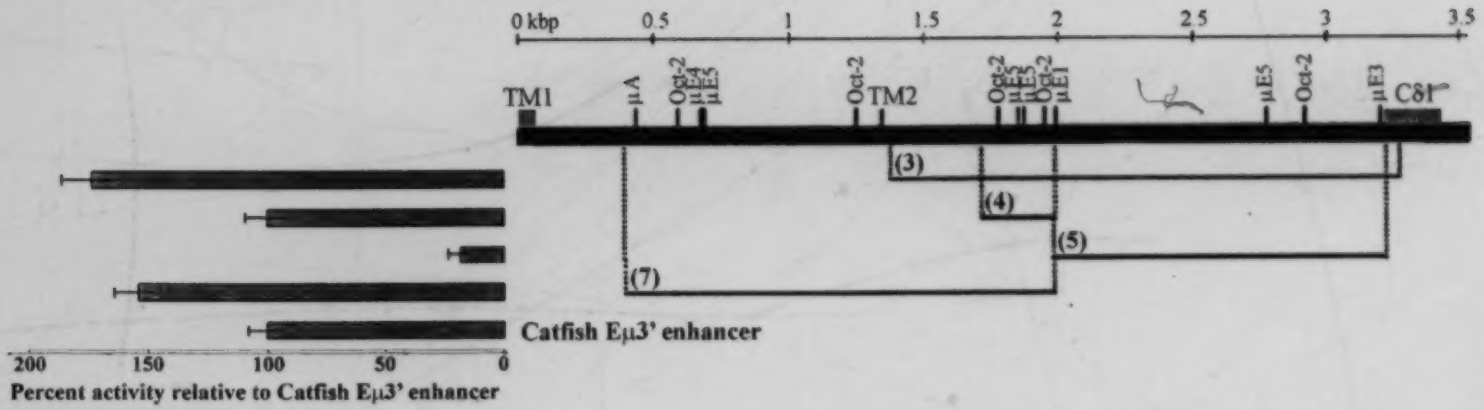

0 kbp
0.5
1
1.5
2
2.5
3
3.5
TM1
μA
Oct-2
μE4
μE5
Oct-2
TM2
Oct-2
μE5
Oct-2
μE1
μE5
Oct-2
μE3
Cδ1
(3)
(4)
(5)
(7)
Catfish Eμ3' enhancer
200
150
100
50
0
Percent activity relative to Catfish Eμ3' enhancer

Figure 14 – The zebrafish Eμ3' enhancer is functional in murine B-cells. The enhancer activities of zebrafish Cμ-Cδ intergenic region fragments (locations indicated relative to scaled map) were tested in the murine J558L myeloma cell line, as described in Fig. 8. Included on the map of the intergenic region are the positions of putative enhancer motifs, using the designations provided in Table 2. Data represents six replicate transfections for each construct performed in the course of two independent experiments. Data are presented as mean percentage of the activity of the catfish Eμ3' enhancer, ± SEM.

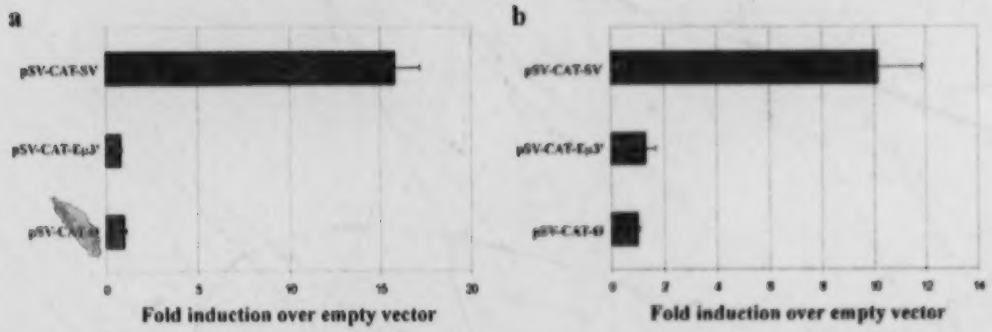


Figure 15 – The zebrafish Eμ3' enhancer is not active in a) fish or b) mammalian T-cell lines. Comparisons were made of the transcriptional activities of chloramphenicol acetyl-transferase (CAT) reporter constructs containing an SV40 promoter alone (pSV-CAT-Ø) or in combination with an SV40 enhancer (pSV-CAT-SV) or Eμ3' fragment 3 (pSV-CAT-Eμ3', see Fig. 8) transiently transfected into the catfish T-cell line 28S.3 (a) or mouse T-cell line EL4 (b). CAT reporter activities were measured after 48h and then corrected and normalized for transfection efficiency using a co-transfected pGL3 luciferase reporter. Enhancer activities are presented as fold induction over the enhancerless control pSV-CAT- Ø, ± SEM. The pSV-CAT-Eμ3' construct is more transcriptionally active than the enhancerless control construct when tested in B-cell lines (1B10: 2-fold induction over enhancerless reporter, J558L: 3.8-fold induction over enhancerless reporter). Three replicate transfections were conducted for each reporter construct.

The zebrafish Cμ-Cδ intron shows no evidence of a remnant transposon

Studies of the catfish *IgH* enhancer indicated the presence of a remnant Tc1-mariner type transposon in the catfish Cμ-Cδ intergenic region (Fig. 13). This transposon was noted as a possible vehicle for evolutionary repositioning of the enhancer since it encompasses some predicted enhancer motifs (Magor et al. 1999). Analysis of the zebrafish Cμ-Cδ intergenic region revealed nothing resembling a remnant transposon in the sequence. In addition, the zebrafish Cμ-Cδ intergenic region is smaller than that of the catfish by approximately one kilobase which corresponds to the size of the catfish transposon. Additionally, there was no evidence of a remnant transposon in Cμ-Cδ sequences from Atlantic salmon (Genbank accession nos. Y12391 and AF141605) and Japanese flounder (Genbank accession no. AB158556).

Discussion:

Eμ3' enhancers among teleost fish share common functional and structural features

The results presented in this study support previous predictions that an Eμ3'-like enhancer in the Cμ-Cδ intergenic is a common feature of the teleost fish (Magor et al. 1994a; Magor et al. 1997). As was seen in the channel catfish *IgH* locus, the zebrafish Cμ-Cδ intergenic region was found to contain a functional enhancer, Eμ3. Like catfish Eμ3', the zebrafish enhancer appears to be organized as a core region with flanking regions that cooperate to mediate transcriptional activation. The core regions of the catfish and zebrafish enhancers both contain a pair of closely spaced E-box motifs, which share a core CANNTG sequence. In each sequence one of these E-boxes is a canonical μE5 box, CAGGTG, which, along with two proximal octamer motifs, was previously shown to act as the core of the catfish enhancer (see Fig. 12a). The second E-box in the catfish Eμ3' was not previously identified and therefore has not been tested for function. Disruption of one or both of the paired E-boxes in the zebrafish Eμ3' core established that they are both critical to the function of the enhancer. A similar closely spaced pair of E-boxes was also found in Cμ-Cδ intergenic sequences from the Atlantic salmon and Japanese flounder (*Paralichthys olivaceus*) (Fig. 12a). In these fish neither of the predicted paired E-boxes are of consensus μE5 sequence (CAGGTG), but one member of each pair has the sequence CACCTG which has been shown to be functional in fish B-cells (Hikima et al. 2004). The predicted Eμ3' regions of these fish have not yet been tested for enhancer function.

In the zebrafish J_H-Cμ1 intron, which corresponds to the location of Eμ in mammals, we also noted the presence of octamer motifs and 3 predicted E-box motifs.

The two closest E-boxes are separated by about 300 bp, and none of the predicted E-boxes are of the sequence CAGGTG or CACCTG. When tested in reporter constructs, the J_H-Cμ1 intron did not display appreciable enhancer activity. Similar observations were made in the catfish Eμ3' enhancer studies. Furthermore, when the J_H-Cμ1 region was inserted into a reporter containing a core-and-flank containing fragment (defined by primers P-102/106) of the zebrafish Eμ3', it did not increase enhancer activity. Despite these results, we cannot exclude the possibility that motifs in the J_H-Cμ intron may interact with the Eμ3' enhancer and contribute to transcription *in vivo*.

Both the zebrafish and catfish Cμ-Cδ intergenic regions were predicted to contain putative MARs. These elements, a common feature of mammalian Eμ enhancers, are a site of attachment of chromatin to the nuclear matrix. In transgenic assays the Eμ MARs have been shown to mediate chromatin accessibility at a distal promoter (Jenuwein et al. 1997; Fernandez et al. 2001) and are required for normal *IgH* expression (Forrester et al. 1994). Two predicted MARs are situated 3' of the zebrafish Eμ3' enhancer core region. Using our reporter assay system, we are unable to determine the effects, if any, of the MAR on zebrafish enhancer activity. A key weakness of our assay system is that an enhancer reporter plasmid cannot reproduce the chromatin configuration at the enhancer in the endogenous *IgH* locus. If the MARs organize or participate in organizing the enhancer-containing region of the zebrafish *IgH* locus in a specific functional configuration *in vivo*, it could alter how factors binding at functional motifs within Eμ3' interact. Within the Atlantic salmon Cμ-Cδ intergenic region a putative 335 bp MAR was identified (bases 14021-14355 of Genbank accession no. Y12391), further indicating that MARs may play a key functional role in endogenous fish Eμ3' enhancers.

Like the catfish Eμ3' enhancer, the zebrafish Eμ3' drove transcription in both mouse and catfish B-cell lines, but had no appreciable activity in T-cell lines. In catfish 1B10 B-cells, the most active zebrafish Eμ3' enhancer construct (construct no. 3, Fig. 8) had approximately 70% of the activity of catfish Eμ3'. In contrast, in murine J558L B-cells the same construct had approximately 170% of the activity of the catfish Eμ3' (Fig. 14). The extent to which inclusion of flanking regions with the core boosted transcriptional activity also varied between the cell lines – the effect appeared more pronounced in catfish B-cells than in mouse B-cells. These findings indicate cell-type specific preferences for particular enhancer motifs, an effect also noted during the characterization of catfish Eμ3'. Because no zebrafish B-cell lines are available, we are unable to determine whether the zebrafish Eμ3' enhancer would have differential activities when tested in zebrafish versus catfish B-cells.

Teleost transcription factors implicated in Eμ3' enhancer function

Several teleost orthologues of mammalian E-protein genes have so far been identified. CFEB-1 and CFEB-2 are two E-protein family members that were recently shown to activate transcription in catfish 1B10 B-cells from a reporter construct containing the catfish enhancer core (Hikima et al. 2004). These factors are related to the mammalian *HEB* gene derived factors. Antibody supershift assays verified the association of CFEB-1 and -2 with the consensus μE5 site and its disruption eliminated the ability of these factors to transactivate when co-expressed in reporter transfected cells. CFEB-1 and -2 were also shown to activate from the non-canonical sequence CACCTG (Hikima et al. 2004) which is the sequence of one of the predicted paired E-

box motifs in both the Atlantic salmon and Japanese flounder (Fig. 12a). *HEB* homologues have not yet been described in zebrafish, however Hikima and colleagues identified a predicted zebrafish *HEB* homologue (Ensembl transcript ID ENSDART00000009938) which grouped closely with CFEB-1 and CFEB-2 of catfish on phylogenetic trees of known vertebrate E-proteins (Hikima et al. 2005a). A tBLASTn search compares a peptide sequence to all six reading frames of nucleic acid sequence in a database. A tBLASTn search of the ENSEMBL zebrafish genome database (www.ensembl.org) using the inferred 705 aa peptide sequence of CFEB-1 (Genbank accession no. AY528668) identified a region on chromosome 7 with significant identity evidenced by several high-scoring hits on a single scaffold which spanned nearly the entire query sequence (see appendix 4). Using the same peptide sequence to search the Japanese Medaka (*Oryzias latipes*) genome database (http://shigen.lab.nig.ac.jp/medaka/genome) also resulted in the identification of homologous sequences with significant alignment scores, especially to the C-terminal portion of the query sequence (data not shown). A single scaffold could not be identified encompassing significant hits across the entire length of the query sequence, which encodes two activation domains and a C-terminal bHLH domain highly conserved among known vertebrate E-proteins. Synteny between the predicted zebrafish and medaka CFEB-1 homologous regions cannot be determined as scaffolds have not yet been physically mapped to particular chromosomes in the *O. latipes* genome sequencing project.

An orthologue of the mammalian *E2A* gene has also been recently cloned from the channel catfish which can interact with the catfish Eμ3' core μE5 site and activate

transcription *in vitro* (Hikima et al. 2005b). In catfish B-cells however the CFEB factors appear to be much more highly expressed (approximately three- to four- fold) and have been suggested to make a greater contribution to transcriptional activation from the Eμ3' enhancer (Hikima et al. 2004; Hikima et al. 2005b). An *E2A* orthologue has also been identified in *Takifugu rubripes* genomic sequence (Hikima et al. 2005a). The transcription factors E12 and E47, which arise from alternate splicing of the mammalian *E2A* gene, would bind the μE5 site in mammalian cells. In zebrafish, transcripts with homology to E12 are known to be expressed (Wulbeck et al. 1994). Hikima and colleagues also identified three predicted zebrafish transcripts with homology to *E2A* (ENSEMBL transcript ID ENSDART00000015857, ENSDART00000018197, ENSDART00000023718) and which group closely with the catfish *E2A* orthologue on phylogenetic trees (Hikima et al. 2005a).

In addition to the core E-box motifs that were shown to be functional, we also identified a number of putative octamer motifs within the zebrafish Eμ3' enhancer. This is consistent with findings from the catfish enhancer studies, where octamer motifs of variant and canonical sequence were shown to be functional and identified as important in mediating transcriptional activation from the catfish Eμ3' enhancer (Magor et al. 1997). In the channel catfish two isoforms of Oct2 (Oct2α, Oct2β), have been identified which arise by alternate splicing of a single gene (Ross et al. 1998). Oct2β has a 34 amino acid C-terminal extension relative to Oct2α. Both isoforms group closely with human and mouse Oct2 on a phylogenetic tree of vertebrate Oct1 and Oct2 amino acid sequences. Catfish Oct2 has been shown to bind the consensus octamer (ATGCAAAT) and a variant octamer (ATGtAAAT) motif, of which there are two copies in the catfish

Eμ3' core region. Catfish Oct2 was also shown to have a preference for driving transcription from the latter (non-canonical) octamer motif (Ross et al. 1998) in both mouse and catfish B-cells. No Oct2 homologue has yet been described in zebrafish. A tBLASTn search of the ENSEMBL zebrafish database using the inferred peptide sequence of Oct2β (480 aa, Genbank accession no. Y12652) resulted in the identification of several high-scoring hits (see appendix 5). Several hits encompassing various regions from residues 73 – 466 of the Oct2β query sequence localized to a single sequence scaffold (Zv4_scaffold1635) which maps to zebrafish chromosome 19. Significant hits encompassing various regions from residues 75 – 392 of the query sequence localized to a second scaffold (Zv4_scaffold1613) also mapped to chromosome 19. These findings support the idea that a homologue of catfish Oct2 is present in the zebrafish genome.

Implications of an Eμ3'-like enhancer as a common feature of teleost fish IgH loci

It has been argued that the Cμ-Cδ intergenic region is an enhancer location that would be unfavorable for the evolution of affinity maturation in teleost fish (Magor et al. 1999). If a similar enhancer location existed prior to the divergence of the teleost fish from the ancestral vertebrate lineage, then the evolution of CSR as seen in higher vertebrate *IgH* loci must have been preceded by a relocation of the main transcriptional enhancer (or at least a functional enhancer) to the J_H-Cμ1 intron or the *de novo* development of an enhancer in this region. Somatic hypermutation was shown to depend on the presence of the Eμ enhancer – an SV40 enhancer was transcriptionally active but unable to drive SH in an *in vitro* assay (Bachl et al. 1998). It is unknown which feature of the enhancer acts to recruit the AID mutator. Recently, several genes targeted for mutation in AID-overexpressing transgenic mice were identified, and the presence of a

common E47 (μE5) binding site proximal to these genes was noted (Kotani et al. 2005). The presence of μE5 motifs has also been implicated in enhancing the rate of somatic hypermutation in a test insert inserted into the variable region of immunoglobulin light chain transgenes (Michael et al. 2003). If the presence of this motif is required to recruit AID and associated factors, the zebrafish and catfish Eμ3' enhancers would satisfy such a requirement in their *IgH* loci. Not only does the Eμ enhancer appear to be required for recruiting the SH machinery, but the rate of mutation has also been shown in an *in vitro* assay to be affected by the distance between the promoter and enhancer – increasing the distance between these elements significantly decreases the mutation rate (Bachl et al. 1998). If somatic hypermutation does occur in fish and these *in vitro* data are relevant, the acquisition of an enhancer in the more promoter-proximal J_H-Cμ1 intron would thus be expected to confer a significant selective advantage.

One method of the evolutionary repositioning of the enhancer that was previously suggested by data from the catfish studies was the transposition of a part of the enhancer to the J_H-Cμ location. In the zebrafish *IgH* locus we saw no evidence of the Tc1 type transposon remnant identified in the catfish Cμ-Cδ intron, nor did we find any evidence of this element in other fish Cμ-Cδ intergenic sequences. The finding that no transposon resides in the zebrafish Cμ-Cδ region or those of other fish indicates that this hypothesis is unlikely to be correct, however it cannot be ruled out that a transposon was involved in such an enhancer "repositioning" event.

With the exception of the Eμ3' enhancer cores, the catfish and zebrafish Cμ-Cδ regions are quite divergent in sequence (Fig. 10a), while the J_H-Cμ1 enhancer regions of mouse and humans are highly conserved (Fig. 10b). It is possible that genetic drift and

balancing selection gave rise to the divergent yet similarly arranged catfish and zebrafish Eμ3' enhancers. A J_H-Cμ1 enhancer could have also arisen in this manner – with the acquisition of functional motifs by random mutation and subsequent selection. The ability of transcriptional regulatory regions to undergo rapid sequence divergence while maintaining their function is well established (reviewed in Tautz 2000). The high conservation and spacing requirements of the mammalian Eμ enhancer suggests that it is an evolutionarily "optimized" regulatory element.

A putative transcriptional enhancer in the zebrafish J_H-Cζ intron

The recently discovered Cζ constant region in the zebrafish *IgH* locus (Danilova et al. 2005) is situated well upstream of the Cμ and Cδ constant regions (Fig. 5). RT-PCR analyses indicated that the Cζ constant region is expressed earlier in development than Cμ. Transcripts of IgZ were shown to always contain sequences from D and J elements (Dζ, Jζ) proximal to the Cζ constant region in the locus (Danilova et al. 2005). IgM transcripts also always contained Cμ-proximal D and J element sequences (Dμ, Jμ). It thus appears that in IgM expressing B-cells the Cζ constant region is spliced from the zebrafish *IgH* locus by a V to DJ recombination event to elements upstream of Cμ (Danilova et al. 2005). The diversity generated by V to DJ recombination to the Dζ and Jζ elements would be limited as there are only two of each of these elements in the locus compared to five each of Dμ and Jμ.

Interestingly, the J_H-Cζ intron may represent a possible location for a transcriptional enhancer, as it contains a number of putative transcription factor binding motifs as was noted when the Cζ region was identified (Danilova et al. 2005). Analysis

of this region using TESS reveals the presence of several putative E-boxes, one of which is of consensus μE5 sequence (CAGGTG). The region also contains predicted binding sites for CBF, PU.1, and Ets-1 located within close proximity to each other although association of these factors with the sites has not been established. A zPicture analysis of the J_H-Cζ region with J_H-Cμ of mouse revealed no significant similarity indicating no apparent relationship between the J_H-Cζ intron and the murine J_H-Cμ region. A similar comparison using the *Xenopus laevis* J_H-Cμ intron gave the same result. However, preliminary analyses of the zebrafish J_H-Cζ region in our lab indicate that it displays enhancer activity *in vitro* (B. Magor, unpublished data).

An alternate scenario of IgH enhancer evolution

It has been argued that the acquisition of a J_H-Cμ intronic enhancer in the *IgH* locus at some point after the divergence of the teleosts was a necessity to allow the evolution of class switching and possibly somatic hypermutation as it occurs in the tetrapods. However, the presence of a putative enhancer in the J_H-Cζ intron suggests another possible route for the emergence of a J_H-Cμ enhancer as seen in higher vertebrates, if a J_H-Cζ enhancer was a feature of the ancestral *IgH* locus prior to the divergence of the teleost fish. It is unknown if catfish possess a Cζ-like constant region because no sequence information is available for the region of the catfish *IgH* locus where it would be predicted to exist. However some evidence suggests that a Cζ-like constant region or remnant thereof is present in several teleosts. Rainbow trout (*Oncorhynchus mykiss*) cDNA sequences encoding IgZ-like heavy chains have been identified (Hansen et al. 2005), as have putative transcripts from carp and salmon

(Danilova et al. 2005). A possible Cζ-like exon cluster located approximately 8 kbp upstream of Cμ in *Takifugu rubripes* genome sequence has also been predicted (Danilova et al. 2005). In an *IgH* locus with a layout like that of the zebrafish, a deletion event in which the region including Cζ to just 5' of Cμ was lost from the locus could reposition a former J_H-Cζ enhancer proximal and 5' of Cμ.

Implications of a J_H-Cζ enhancer for affinity maturation in fish

Fish clearly express AID, and although they show poor affinity maturation it appears that stimulated catfish B-cells expressing AID accumulate mutations (Saunders and Magor, unpublished observations). Furthermore, preliminary evidence suggest that catfish AID is able to drive CSR in mammalian B-cells (Muramatsu, Magor and colleagues, unpublished observations). If B-cells expressing IgZ in zebrafish are capable of undergoing affinity maturation after activation, a predicted feature of a more V_H promoter-proximal J_H-Cζ enhancer would be higher rates of somatic hypermutation of the rearranged VDJ exon in IgZ-expressing cells compared to IgM expressing cells. Furthermore, mutations that accumulate within the rearranged V element in B-cells expressing IgZ could possibly be retained after a subsequent switch to IgM by a rearranged-V to DμJμ recombination event, although no evidence for this occurrence has yet been described. IgZ is expressed early in development and it is unknown whether it has biological effector functions distinct from IgM antibodies or what the nature of these effector functions are. Given the early appearance of IgZ-expressing B-cells in zebrafish development, it has been suggested that IgZ antibodies may have a specialized role in defense against insults commonly encountered during early maturation (Danilova et al.

2005). If a V-D-J-Cζ-D-J-Cμ-Cδ like locus (with a J_H-Cζ enhancer) actually represents the ancestral locus arrangement prior to the divergence of the teleosts, then the development of a J_H-Cμ intronic enhancer may not have been a limiting step in the evolution of affinity maturation in the *IgH* locus as has been suggested (Magor et al. 1999).

References:

Arakawa, H. and J. M. Buerstedde (2004). "Immunoglobulin gene conversion: insights from bursal B cells and the DT40 cell line." Dev Dyn **229**(3): 458-64.

Arakawa, H., J. Hauschild and J. M. Buerstedde (2002). "Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion." Science **295**(5558): 1301-6.

Bachl, J., C. Olsson, N. Chitkara and M. Wabl (1998). "The Ig mutator is dependent on the presence, position, and orientation of the large intron enhancer." Proc Natl Acad Sci U S A **95**(5): 2396-9.

Banerji, J., L. Olson and W. Schaffner (1983). "A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes." Cell **33**(3): 729-40.

Beckmann, H., L. K. Su and T. Kadesch (1990). "TFE3: a helix-loop-helix protein that activates transcription through the immunoglobulin enhancer muE3 motif." Genes Dev **4**(2): 167-79.

Bengten, E., S. M. Quiniou, T. B. Stuge, T. Katagiri, N. W. Miller, L. W. Clem, G. W. Warr and M. Wilson (2002). "The IgH locus of the channel catfish, Ictalurus punctatus, contains multiple constant region gene sequences: different genes encode heavy chains of membrane and secreted IgD." J Immunol **169**(5): 2488-97.

Betz, A. G., M. S. Neuberger and C. Milstein (1993). "Discriminating intrinsic and antigen-selected mutational hotspots in immunoglobulin V genes." Immunol Today **14**(8): 405-11.

Binz, S. K., A. M. Sheehan and M. S. Wold (2004). "Replication protein A phosphorylation and the cellular response to DNA damage." DNA Repair (Amst) **3**(8-9): 1015-24.

Blackwood, E. M. and J. T. Kadonaga (1998). "Going the distance: a current view of enhancer action." Science **281**(5373): 60-3.

Cain, K. D., D. R. Jones and R. L. Raison (2002). "Antibody-antigen kinetics following immunization of rainbow trout (Oncorhynchus mykiss) with a T-cell dependent antigen." Dev Comp Immunol **26**(2): 181-90.

Calame, K. and R. Sen (2004). Molecular Biology of B-cells. Amsterdam; Boston, Elsevier: 83-100.

Carey, M. and S. T. Smale (2000). Transcriptional regulation in eukaryotes: concepts, strategies, and techniques. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.

Chaudhuri, J. and F. W. Alt (2004). "Class-switch recombination: interplay of transcription, DNA deamination and DNA repair." Nat Rev Immunol **4**(7): 541-52.

Chaudhuri, J., C. Khuong and F. W. Alt (2004). "Replication protein A interacts with AID to promote deamination of somatic hypermutation targets." Nature **430**(7003): 992-8.

Chen, H. M., P. Zhang, M. T. Voso, S. Hohaus, D. A. Gonzalez, C. K. Glass, D. E. Zhang and D. G. Tenen (1995). "Neutrophils and monocytes express high levels of PU.1 (Spi-1) but not Spi-B." Blood **85**(10): 2918-28.

Choi, J. K., C. P. Shen, H. S. Radomska, L. A. Eckhardt and T. Kadesch (1996). "E47 activates the Ig-heavy chain and TdT loci in non-B cells." EMBO J **15**(18): 5014-21.

Chowdhury, D. and R. Sen (2004). "Regulation of immunoglobulin heavy-chain gene rearrangements." Immunol Rev **200**: 182-96.

Cioffi, C. C., D. L. Middleton, M. R. Wilson, N. W. Miller, L. W. Clem and G. W. Warr (2001). "An IgH enhancer that drives transcription through basic helix-loop-helix and Oct transcription factor binding motifs. Functional analysis of the E(mu)3' enhancer of the catfish." J Biol Chem **276**(30): 27825-30.

Cooper, C., D. Johnson, C. Roman, N. Avitahl, P. Tucker and K. Calame (1992). "The C/EBP family of transcriptional activators is functionally important for Ig VH promoter activity in vivo and in vitro." J Immunol **149**(10): 3225-31.

Dang, W., B. S. Nikolajczyk and R. Sen (1998). "Exploring functional redundancy in the immunoglobulin mu heavy-chain gene enhancer." Mol Cell Biol **18**(11): 6870-8.

Danilova, N., J. Bussmann, K. Jekosch and L. A. Steiner (2005). "The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z." Nat Immunol **6**(3): 295-302.

Descombes, P. and U. Schibler (1991). "A liver-enriched transcriptional activator protein, LAP, and a transcriptional inhibitory protein, LIP, are translated from the same mRNA." Cell **67**(3): 569-79.

Di Noia, J. and M. S. Neuberger (2002). "Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase." Nature **419**(6902): 43-8.

Du Pasquier, L. (2001). "The immune system of invertebrates and vertebrates." Comp Biochem Physiol B Biochem Mol Biol **129**(1): 1-15.

Du Pasquier, L., M. Wilson, A. S. Greenberg and M. F. Flajnik (1998). "Somatic mutation in ectothermic vertebrates: musings on selection and origins." Curr Top Microbiol Immunol **229**: 199-216.

Dunnick, W. A., J. Shi, K. A. Graves and J. T. Collins (2005). "The 3' end of the heavy chain constant region locus enhances germline transcription and switch recombination of the four {gamma} genes." J Exp Med **201**(9): 1459-66.

Eason, D. D., J. P. Cannon, R. N. Haire, J. P. Rast, D. A. Ostrov and G. W. Litman (2004). "Mechanisms of antigen receptor evolution." Semin Immunol **16**(4): 215-26.

Engel, I. and C. Murre (2001). "The function of E- and Id proteins in lymphocyte development." Nat Rev Immunol **1**(3): 193-9.

Erman, B., M. Cortes, B. S. Nikolajczyk, N. A. Speck and R. Sen (1998). "ETS-core binding factor: a common composite motif in antigen receptor gene enhancers." Mol Cell Biol **18**(3): 1322-30.

Ernst, P. and S. T. Smale (1995). "Combinatorial regulation of transcription II: The immunoglobulin mu heavy chain gene." Immunity **2**(5): 427-38.

Fernandez, L. A., M. Winkler and R. Grosschedl (2001). "Matrix attachment region-dependent function of the immunoglobulin mu enhancer involves histone acetylation at a distance without changes in enhancer occupancy." Mol Cell Biol **21**(1): 196-208.

Flajñik, M. F. (2002). "Comparative analyses of immunoglobulin genes: surprises and portents." Nat Rev Immunol **2**(9): 688-98.

Forrester, W. C., C. van Genderen, T. Jenuwein and R. Grosschedl (1994). "Dependence of enhancer-mediated transcription of the immunoglobulin mu gene on nuclear matrix attachment regions." Science **265**(5176): 1221-5.

Frazer, J. K. and J. D. Capra (1999). Immunoglobulins: Structure and Function. Fundamental Immunology. W. E. Paul. Philadelphia, Lippincott-Raven: 37-74.

Frisch, M., K. Frech, A. Klingenhoff, K. Cartharius, I. Liebich and T. Werner (2002). "In silico prediction of scaffold/matrix attachment regions in large genomic sequences." Genome Res **12**(2): 349-54.

Galson, D. L., J. O. Hensold, T. R. Bishop, M. Schalling, A. D. D'Andrea, C. Jones, P. E. Auron and D. E. Housman (1993). "Mouse beta-globin DNA-binding protein B1 is identical to a proto-oncogene, the transcription factor Spi-1/PU.1, and is restricted in expression to hematopoietic cells and the testis." Mol Cell Biol **13**(5): 2929-41.

Ghosh, D. (1990). "A relational database of transcription factors." Nucleic Acids Res **18**(7): 1749-56.

Gillies, S. D., S. L. Morrison, V. T. Oi and S. Tonegawa (1983). "A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene." Cell **33**(3): 717-28.

Grosschedl, R., K. Giese and J. Pagel (1994). "HMG domain proteins: architectural elements in the assembly of nucleoprotein structures." Trends Genet **10**(3): 94-100.

Gstaiger, M., L. Knoepfel, O. Georgiev, W. Schaffner and C. M. Hovens (1995). "A B-cell coactivator of octamer-binding transcription factors." Nature **373**(6512): 360-2.

Hansen, J. D., E. D. Landis and R. B. Phillips (2005). "Discovery of a unique Ig heavy-chain isotype (IgT) in rainbow trout: Implications for a distinctive B cell developmental pathway in teleost fish." Proc Natl Acad Sci U S A **102**(19): 6919-24.

Hatada, E. N., S. Chen-Kiang and C. Scheidereit (2000). "Interaction and functional interference of C/EBPbeta with octamer factors in immunoglobulin gene transcription." Eur J Immunol **30**(1): 174-84.

Herrscher, R. F., M. H. Kaplan, D. L. Lelsz, C. Das, R. Scheuermann and P. W. Tucker (1995). "The immunoglobulin heavy-chain matrix-associating regions are bound by Bright: a B cell-specific trans-activator that describes a new DNA-binding protein family." Genes Dev **9**(24): 3067-82.

Hikima, J., C. C. Cioffi, D. L. Middleton, M. R. Wilson, N. W. Miller, L. W. Clem and G. W. Warr (2004). "Evolution of transcriptional control of the IgH locus: characterization, expression, and function of TF12/HEB homologs of the catfish." J Immunol **173**(9): 5476-84.

Hikima, J. I., M. L. Lennard, M. R. Wilson, N. W. Miller, L. W. Clem and G. W. Warr (2005a). "Evolution of vertebrate E-protein transcription factors: Comparative analysis of the E-protein gene family in Takifugu rubripes and humans." Physiol Genomics.

Hikima, J. I., D. L. Middleton, M. R. Wilson, N. W. Miller, L. W. Clem and G. W. Warr (2005b). "Regulation of immunoglobulin gene transcription in a teleost fish: identification, expression and functional properties of E2A in the channel catfish." Immunogenetics.

Hordvik, I., J. Thevarajan, I. Samdal, N. Bastani and B. Krossoy (1999). "Molecular cloning and phylogenetic analysis of the Atlantic salmon immunoglobulin D gene." Scand J Immunol **50**(2): 202-10.

Imai, K., G. Slupphaug, W. I. Lee, P. Revy, S. Nonoyama, N. Catalan, L. Yel, M. Forveille, B. Kavli, H. E. Krokan, H. D. Ochs, A. Fischer and A. Durandy (2003). "Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination." Nat Immunol **4**(10): 1023-8.

Jenuwein, T., W. C. Forrester, L. A. Fernandez-Herrero, G. Laible, M. Dull and R. Grosschedl (1997). "Extension of chromatin accessibility by nuclear matrix attachment regions." Nature **385**(6613): 269-72.

Jenuwein, T. and R. Grosschedl (1991). "Complex pattern of immunoglobulin mu gene expression in normal and transgenic mice: nonoverlapping regulatory sequences govern distinct tissue specificities." Genes Dev **5**(6): 932-43.

Kaattari, S. L., H. L. Zhang, I. W. Khor, I. M. Kaattari and D. A. Shapiro (2002). "Affinity maturation in trout: clonal dominance of high affinity antibodies late in the immune response." Dev Comp Immunol **26**(2): 191-200.

Kee, B. L., M. W. Quong and C. Murre (2000). "E2A proteins: essential regulators at multiple stages of B-cell development." Immunol Rev **175**: 138-49.

Koop, B. F., J. E. Richards, T. D. Durfee, J. Bansberg, J. Wells, A. C. Gilliam, H. L. Chen, A. Clausell, P. W. Tucker and F. R. Blattner (1996). "Analysis and comparison of the mouse and human immunoglobulin heavy chain JH-Cmu-Cdelta locus." Mol Phylogenet Evol **5**(1): 33-49.

Kotani, A., I. M. Okazaki, M. Muramatsu, K. Kinoshita, N. A. Begum, T. Nakajima, H. Saito and T. Honjo (2005). "A target selection of somatic hypermutations is regulated similarly between T and B cells upon activation-induced cytidine deaminase expression." Proc Natl Acad Sci U S A **102**(12): 4506-11.

Kottmann, A. H., B. Zevnik, M. Welte, P. J. Nielsen and G. Kohler (1994). "A second promoter and enhancer element within the immunoglobulin heavy chain locus." Eur J Immunol **24**(4): 817-21.

Landschulz, W. H., P. F. Johnson and S. L. McKnight (1988). "The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins." Science **240**(4860): 1759-64.

Lekstrom-Himes, J. and K. G. Xanthopoulos (1998). "Biological role of the CCAAT/enhancer-binding protein family of transcription factors." J Biol Chem **273**(44): 28545-8.

Liebich, I., J. Bode, M. Frisch and E. Wingender (2002). "S/MARt DB: a database on scaffold/matrix attached regions." Nucleic Acids Res **30**(1): 372-4.

Litman, G. W., M. F. Flajnik and G. W. Warr (2004). Molecular Biology of B-cells. T. Honjo, F. W. Alt and M. S. Neuberger. Amsterdam; Boston, Elsevier: 417-432.

Loots, G. G. and I. Ovcharenko (2004). "rVISTA 2.0: evolutionary analysis of transcription factor binding sites." Nucleic Acids Res **32**(Web Server issue): W217-21.

Luo, Y. and R. G. Roeder (1995). "Cloning, functional characterization, and mechanism of action of the B-cell-specific transcriptional coactivator OCA-B." Mol Cell Biol **15**(8): 4115-24.

Lutzko, C., D. Senadheera, D. Skelton, D. Petersen and D. B. Kohn (2003). "Lentivirus vectors incorporating the immunoglobulin heavy chain enhancer and matrix attachment regions provide position-independent expression in B lymphocytes." J Virol **77**(13): 7341-51.

Magor, B. G., D. A. Ross, D. L. Middleton and G. W. Warr (1997). "Functional motifs in the IgH enhancer of the channel catfish." Immunogenetics **46**(3): 192-8.

Magor, B. G., D. A. Ross, L. Pilstrom and G. W. Warr (1999). "Transcriptional enhancers and the evolution of the IgH locus." Immunol Today **20**(1): 13-7.

Magor, B. G., M. R. Wilson, N. W. Miller, L. W. Clem, D. L. Middleton and G. W. Warr (1994a). "An Ig heavy chain enhancer of the channel catfish Ictalurus punctatus: evolutionary conservation of function but not structure." J Immunol **153**(12): 5556-63.

Magor, B. G., M. R. Wilson, N. W. Miller, L. W. Clem, D. L. Middleton and G. W. Warr (1994b). "An Ig heavy chain enhancer of the channel catfish Ictalurus punctatus: evolutionary conservation of function but not structure." J Immunol **153**(12): 5556-63.

Mamane, Y., C. Heylbroeck, P. Genin, M. Algarte, M. J. Servant, C. LePage, C. DeLuca, H. Kwon, R. Lin and J. Hiscott (1999). "Interferon regulatory factors: the next generation." Gene **237**(1): 1-14.

Manis, J. P., M. Tian and F. W. Alt (2002). "Mechanism and control of class-switch recombination." Trends Immunol **23**(1): 31-9.

Marecki, S., K. M. McCarthy and B. S. Nikolajczyk (2004). "PU.1 as a chromatin accessibility factor for immunoglobulin genes." Mol Immunol **40**(10): 723-31.

Matthias, P. (1998). "Lymphoid-specific transcription mediated by the conserved octamer site: who is doing what?" Semin Immunol **10**(2): 155-63.

Max, E. E. (1999). Immunoglobulins: Molecular Genetics. Fundamental immunology. W. E. Paul. Philadelphia, Lippincott-Raven: 111-182.

McCarthy, K. M., D. McDevit, A. Andreucci, R. Reeves and B. S. Nikolajczyk (2003). "HMGA1 co-activates transcription in B cells through indirect association with DNA." J Biol Chem **278**(43): 42106-14.

McKercher, S. R., B. E. Torbett, K. L. Anderson, G. W. Henkel, D. J. Vestal, H. Baribault, M. Klemsz, A. J. Feeney, G. E. Wu, C. J. Paige and R. A. Maki (1996). "Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities." EMBO J **15**(20): 5647-58.

Michael, N., H. M. Shen, S. Longerich, N. Kim, A. Longacre and U. Storb (2003). "The E box motif CAGGTG enhances somatic hypermutation without enhancing transcription." Immunity **19**(2): 235-42.

Miller, N. W., M. A. Rycyzyn, M. R. Wilson, G. W. Warr, J. P. Naftel and L. W. Clem (1994). "Development and characterization of channel catfish long term B cell lines." J Immunol **152**(5): 2180-9.

Munshi, N., M. Merika, J. Yie, K. Senger, G. Chen and D. Thanos (1998). "Acetylation of HMG I(Y) by CBP turns off IFN beta expression by disrupting the enhanceosome." Mol Cell **2**(4): 457-67.

Muramatsu, M., K. Kinoshita, S. Fagarasan, S. Yamada, Y. Shinkai and T. Honjo (2000). "Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme." Cell **102**(5): 553-63.

Nagulapalli, S., J. M. Pongubala and M. L. Atchison (1995). "Multiple proteins physically interact with PU.1. Transcriptional synergy with NF-IL6 beta (C/EBP delta, CRP3)." J Immunol **155**(9): 4330-8.

Nelsen, B., G. Tian, B. Erman, J. Gregoire, R. Maki, B. Graves and R. Sen (1993). "Regulation of lymphoid-specific immunoglobulin mu heavy chain gene enhancer by ETS-domain proteins." Science **261**(5117): 82-6.

Nikolajczyk, B. S., B. Nelsen and R. Sen (1996). "Precise alignment of sites required for mu enhancer activation in B cells." Mol Cell Biol **16**(8): 4544-54.

Nikolajczyk, B. S., J. A. Sanchez and R. Sen (1999). "ETS protein-dependent accessibility changes at the immunoglobulin mu heavy chain enhancer." Immunity **11**(1): 11-20.

Ovcharenko, I., G. G. Loots, R. C. Hardison, W. Miller and L. Stubbs (2004). "zPicture: dynamic alignment and visualization tool for analyzing conservation profiles." Genome Res **14**(3): 472-7.

Parkin, S. E., M. Baer, T. D. Copeland, R. C. Schwartz and P. F. Johnson (2002). "Regulation of CCAAT/enhancer-binding protein (C/EBP) activator proteins by heterodimerization with C/EBPgamma (Ig/EBP)." J Biol Chem **277**(26): 23563-72.

Petsko, G. A. (2002). Modeling structure from sequence. Mississauga, Wiley Interscience.

Quong, M. W., W. J. Romanow and C. Murre (2002). "E protein function in lymphocyte development." Annu Rev Immunol **20**: 301-22.

Rada, C., G. T. Williams, H. Nilsen, D. E. Barnes, T. Lindahl and M. S. Neuberger (2002). "Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice." Curr Biol **12**(20): 1748-55.

Reeves, R. and L. Beckerbauer (2001). "HMGI/Y proteins: flexible regulators of transcription and chromatin structure." Biochim Biophys Acta **1519**(1-2): 13-29.

Revy, P., T. Muto, Y. Levy, F. Geissmann, A. Plebani, O. Sanal, N. Catalan, M. Forveille, R. Dufourcq-Labelouse, A. Gennery, I. Tezcan, F. Ersoy, H. Kayserili, A. G. Ugazio, N. Brousse, M. Muramatsu, L. D. Notarangelo, K. Kinoshita, T. Honjo, A. Fischer and A. Durandy (2000). "Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2)." Cell **102**(5): 565-75.

Rogozin, I. B. and M. Diaz (2004). "Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process." J Immunol **172**(6): 3382-4.

Roman, C., J. S. Platero, J. Shuman and K. Calame (1990). "Ig/EBP-1: a ubiquitously expressed immunoglobulin enhancer binding protein that is similar to C/EBP and heterodimerizes with C/EBP." Genes Dev **4**(8): 1404-15.

Ross, D. A., B. G. Magor, D. L. Middleton, M. R. Wilson, N. W. Miller, L. W. Clem and G. W. Warr (1998). "Characterization of Oct2 from the channel catfish: functional preference for a variant octamer motif." J Immunol **160**(8): 3874-82.

Satterthwaite, A. B. and O. N. Witte (2000). "The role of Bruton's tyrosine kinase in B-cell development and function: a genetic perspective." Immunol Rev **175**: 120-7.

Saunders, H. L. and B. G. Magor (2004). "Cloning and expression of the AID gene in the channel catfish." Dev Comp Immunol **28**(7-8): 657-63.

Schwartz, S., W. J. Kent, A. Smit, Z. Zhang, R. Baertsch, R. C. Hardison, D. Haussler and W. Miller (2003). "Human-mouse alignments with BLASTZ." Genome Res **13**(1): 103-7.

Shen, H. M., A. Peters, B. Baron, X. Zhu and U. Storb (1998). "Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes." Science **280**(5370): 1750-2.

Smale, S. T. and J. T. Kadonaga (2003). "The RNA polymerase II core promoter." Annu Rev Biochem **72**: 449-79.

Srisapoome, P., T. Ohira, I. Hirono and T. Aoki (2004). "Genes of the constant regions of functional immunoglobulin heavy chain of Japanese flounder, Paralichthys olivaceus." Immunogenetics **56**(4): 292-300.

Stief, A., D. M. Winter, W. H. Stratling and A. E. Sippel (1989). "A nuclear DNA attachment element mediates elevated and position-independent gene activity." Nature **341**(6240): 343-5.

Strubin, M., J. W. Newell and P. Matthias (1995). "OBF-1, a novel B cell-specific coactivator that stimulates immunoglobulin promoter activity through association with octamer-binding proteins." Cell **80**(3): 497-506.

Tautz, D. (2000). "Evolution of transcriptional regulation." Curr Opin Genet Dev **10**(5): 575-9.

Thanos, D. and T. Maniatis (1995). "Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome." Cell **83**(7): 1091-100.

Wang, Z., A. Goldstein, R. T. Zong, D. Lin, E. J. Neufeld, R. H. Scheuermann and P. W. Tucker (1999). "Cux/CDP homeoprotein is a component of NF-muNR and represses the immunoglobulin heavy chain intronic enhancer by antagonizing the bright transcription activator." Mol Cell Biol **19**(1): 284-95.

Warr, G. W. (1995). "The immunoglobulin genes of fish." Dev Comp Immunol **19**(1): 1-12.

Webb, C. F., Y. Yamashita, N. Ayers, S. Evetts, Y. Paulin, M. E. Conley and E. A. Smith (2000). "The transcription factor Bright associates with Bruton's tyrosine kinase, the defective protein in immunodeficiency disease." J Immunol **165**(12): 6956-65.

Wilsker, D., A. Patsialou, P. B. Dallas and E. Moran (2002). "ARID proteins: a diverse family of DNA binding proteins implicated in the control of cell growth, differentiation, and development." Cell Growth Differ **13**(3): 95-106.

Wilson, M. R., H. Zhou, E. Bengten, L. W. Clem, T. B. Stuge, G. W. Warr and N. W. Miller (1998). "T-cell receptors in channel catfish: structure and expression of TCR alpha and beta genes." Mol Immunol **35**(9): 545-57.

Wulbeck, C., C. Fromental-Ramain and J. A. Campos-Ortega (1994). "The HLH domain of a zebrafish HE12 homologue can partially substitute for functions of the HLH domain of Drosophila DAUGHTERLESS." Mech Dev **46**(2): 73-85.

Yie, J., M. Merika, N. Munshi, G. Chen and D. Thanos (1999). "The role of HMG I(Y) in the assembly and function of the IFN-beta enhanceosome." EMBO J **18**(11): 3074-89.

Yu, K., F. Chedin, C. L. Hsieh, T. E. Wilson and M. R. Lieber (2003). "R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells." Nat Immunol **4**(5): 442-51.

Yuan, D., T. Dang, J. Hawley, T. Jenuwein and R. Grosschedl (1995). "Role of the OCTA site in regulation of IgH chain gene transcription during B cell activation." Int Immunol **7**(8): 1163-72.

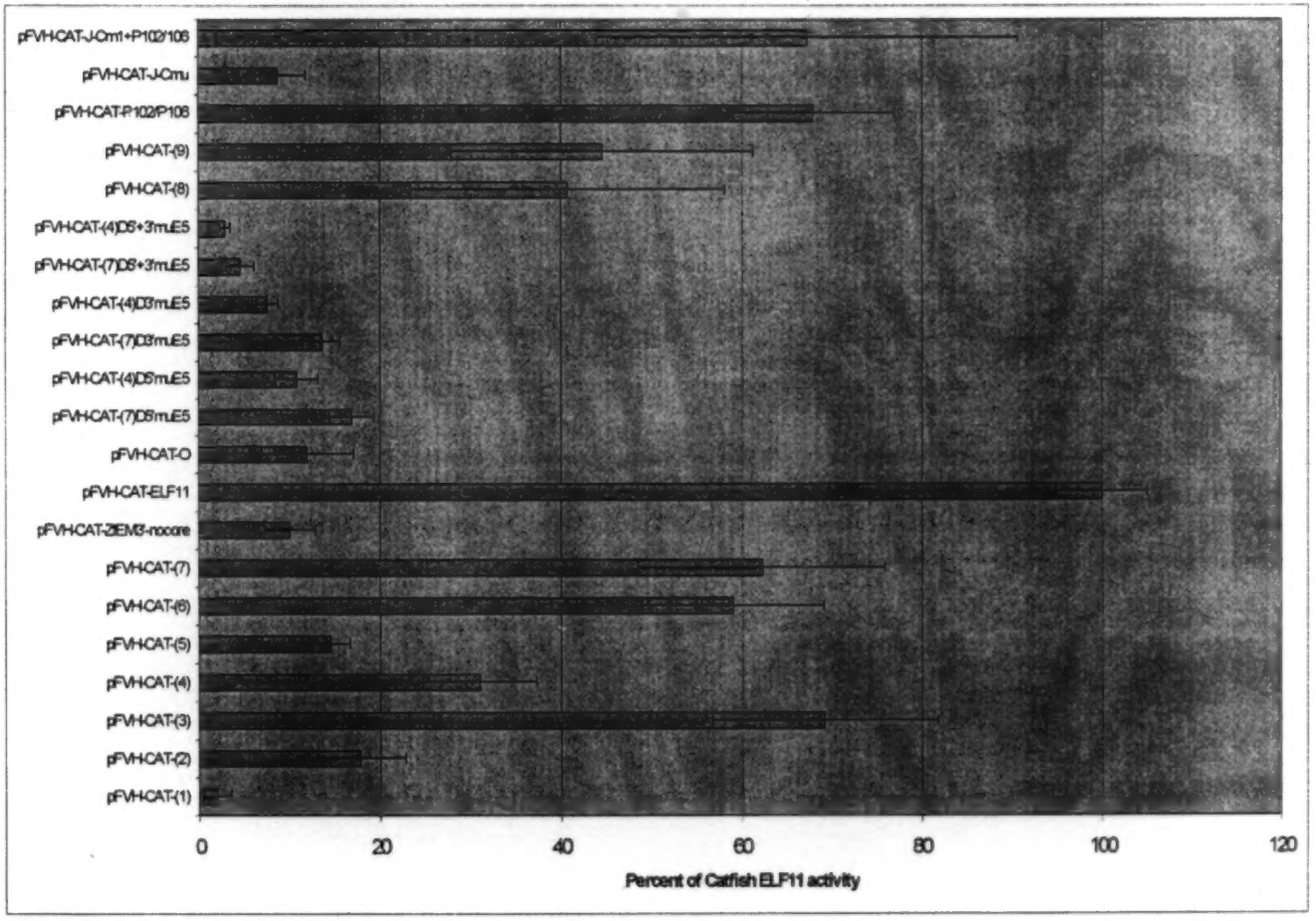
Zarrin, A. A., F. W. Alt, J. Chaudhuri, N. Stokes, D. Kaushal, L. Du Pasquier and M. Tian (2004). "An evolutionarily conserved target motif for immunoglobulin class-switch recombination." Nat Immunol **5**(12): 1275-81.

Zhang, D. E., P. Zhang, N. D. Wang, C. J. Hetherington, G. J. Darlington and D. G. Tenen (1997). "Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice." Proc Natl Acad Sci U S A **94**(2): 569-74.

Zhao, Y., Q. Pan-Hammarstrom, Z. Zhao and L. Hammarstrom (2005). "Identification of the activation-induced cytidine deaminase gene from zebrafish: an evolutionary analysis." Dev Comp Immunol **29**(1): 61-71.

Zwilling, S., A. Annweiler and T. Wirth (1994). "The POU domains of the Oct1 and Oct2 transcription factors mediate specific interaction with TBP." Nucleic Acids Res **22**(9): 1655-62.

Appendix 1 – Compiled 1B10 transfection data expressed as mean percent of Catfish Eμ3' enhancer activity ± 95% confidence intervals (x ± t*SE)



Appendix 2 – *IgH* enhancer motifs from the TFD database (Ghosh 1990) included in motif searches

ATGCAAAT	Oct-2
CAGGTGGC	E-box CS
TGCCACATGA	uE3
AAGATGGC	uE1
CAGCTGGC	uE2
CAGGTGGT	uE4
CCACATGAC	CuE3.4
TGCAGGTGT	CuE5
AACACCTG	mu-E5
TATTTGGGGAA	mu-B
RCAGNTG	E2A
SMGGAWGY	Ets-1
RCAGGTG	E47
RCANNTGTG	E-box
AACACCTG	mu-E5
TGCAGGTGT	CuE5

Appendix 3 - Raw transfection data

1B10

SET1

construct		avg.lucif-mock	CAT-mock	(*)=exclude	CAT/LUC
(4)	A1	5.10	7045		1381
"	A2	7.12	10236		1437
"	A3	7.25	9689		1337
(6)	B1	3.57	10666		2984
"	B2	3.70	9125		2469
"	B3	2.43	6114		2520
(7)	C1	3.13	3983		1271
"	C2	3.62	4757		1316
"	C3	6.28	10696		1702
no core	D1	6.36	2666		419
"	D2	5.37	1845		344
"	D3	0.05	-14	*	-273
ELF11	E1	6.31	32177		5101
"	E2	6.10	32887		5393
"	E3	2.48	18832		7594
empty.vec.	F1	7.77	2563		330
"	F2	9.69	2460		254
"	F3	4.66	834		179
(7)D5'muE5	G1	8.40	7460		888
"	G2	6.13	4998		816
"	G3	13.91	9941		715
(4)D5'muE5	H1	9.77	5601		573
"	H2	13.11	7357		561
"	H3	11.12	5274		474
(7)D3'muE5	I1	7.71	5481		711
"	I2	7.40	6399		864
"	I3	9.82	5967		608
(4)D3'muE5	J1	7.60	2865		377
"	J2	5.35	2838		530
"	J3	3.66	1350		369
(7)D5'+3'muE5	K1	10.49	2592		247
"	K2	1.52	91		60
"	K3	5.93	1101		186
(4)D5'+3'muE5	L1	7.98	1031		129
"	L2	8.59	1403		163
"	L3	8.97	1250		139
(8)	M1	6.47	6347		980
"	M2	7.18	6465		900
"	M3	6.05	5699		943
(9)	N1	5.85	5726		978
"	N2	5.26	6957		1323
"	N3	3.77	4757		1262
(2)	O1	5.94	3613		608

(2)	O2	6.40	4914		768
"	O3	3.33	2546		764
	Mock	0.00	0		

SET2

construct		avg.lucif-mock	CAT-mock	(*)=exclude	CAT/LUC
(4)	A1	2.53	4208		1666
"	A2	4.39	7166		1631
"	A3	1.62	3417		2114
(7)	C1	2.08	4341		2091
"	C2	1.94	4036		2080
"	C3	1.19	1978		1659
no core	D1	2.18	1141		523
"	D2	2.79	1458		523
"	D3	0.96	563		587
ELF11	E1	2.32	11054		4769
"	E2	2.78	10949		3933
"	E3	1.71	5356		3136
empty vec.	F1	3.26	870		267
"	F2	1.17	550		469
"	F3	0.00	452	*	112960
(7)D5'muE5	G1	2.70	2363		876
"	G2	4.09	2863		700
"	G3	2.61	1857		710
(7)D3'muE5	I1	2.72	1760		647
"	I2	2.64	1563		591
"	I3	0.05	447	*	9308
(7)D5'+3'muE5	K1	2.73	576		211
"	K2	0.30	699	*	2345
"	K3	1.96	460		235
(8)	M1	0.64	330		515
"	M2	0.52	362		691
"	M3	1.70	708		417
(9)	N1	0.25	92	*	361
"	N2	1.06	805		756
"	N3	0.53	387		733
	Mock	0.00	0		

SET3

construct		avg.lucif-mock	CAT-mock	(*)=exclude	CAT/LUC
(3)	A1	0.39	527		1352
"	A2	8.04	15297		1903
"	A3	-0.03	-42	*	1291
(4)	B1	0.00	10	*	-2024
"	B2	23.13	16585		717
"	B3	10.40	7720		742
(5)	C1	3.60	1537		427
"	C2	3.53	1361		385
"	C3	17.64	9142		518

(6)	D1	11.76	17092		1453
"	D2	8.81	10992		1248
"	D3	11.25	19524		1736
(7)	E1	13.62	12321		904
"	E2	11.74	12198		1039
"	E3	0.09	27	*	311
J-Cm+P102/106	F1	1.56	4012		2565
"	F2	1.10	3257		2950
"	F3	1.49	3326		2232
no core	G1	9.70	3162		326
"	G2	11.22	4243		378
"	G3	8.29	2766		334
ELF11	H1	8.27	26655		3224
"	H2	1.33	6010		4512
"	H3	8.89	33504		3770
empty vec.	I1	11.34	2839		250
"	I2	11.14	3004		270
"	I3	11.05	3187		289
(7)D5'muE5	J1	12.92	9469		733
"	J2	11.73	8530		727
"	J3	10.55	6644		630
(4)D5'muE5	K1	12.04	6414		533
"	K2	15.89	7278		458
"	K3	9.36	4507		482
(7)D3'muE5	L1	10.83	6650		614
"	L2	2.27	909		400
"	L3	10.82	6269		579
(4)D3'muE5	M1	11.92	3531		296
"	M2	14.02	4646		331
"	M3	13.10	4155		317
(7)D5'+3'muE5	N1	11.17	2089		187
"	N2	1.12	310		277
"	N3	5.99	1167		195
(4)D5'+3'muE5	O1	13.50	1692		125
"	O2	15.11	2018		134
"	O3	9.52	1206		127
(8)	P1	9.38	7860		838
"	P2	10.76	8741		812
"	P3	0.15	128	*	877
(9)	Q1	6.66	6520		979
"	Q2	6.25	4765		763
"	Q3	5.96	5624		944
	Mock	0.00	0		

SET4

construct		avg.lucif-mock	CAT-mock	(*)=exclude	CAT/LUC
(7)	A1	1.58	5634		3569
"	A2	1.69	6707		3972

(7)	A3	1.66	7618		4588
(6)	B1	1.17	5246		4467
"	B2	-0.05	372	*	
"	B3	1.09	4204		3841
(5)	C1	0.88	759		864
"	C2	1.30	1153		884
"	C3	1.77	1558		879
(6)	D1	0.92	3650		3948
"	D2	0.80	3174		3965
"	D3	1.13	4168		3700
(4)	E1	1.40	2312		1655
"	E2	0.90	1632		1808
"	E3	1.04	1950		1871
(5)	F1	0.69	722		1045
"	F2	1.19	958		802
"	F3	1.04	957		918
P199/200	G1	3.12	2257		723
"	G2	1.73	1569		905
"	G3	1.73	1712		991
(1)	H1	1.90	435		229
"	H2	1.86	391		210
"	H3	1.77	321		181
(2)	I1	1.53	1757		1151
"	I2	1.59	1569		989
"	I3	1.72	1605		934
(3)	J1	1.62	6031		3722
"	J2	1.80	6890		3835
"	J3	0.00	61	*	
J-Cmu	K1	0.61	273		451
"	K2	0.43	381		884
"	K3	0.80	526		659
P102/106	L1	1.96	5790		2959
"	L2	0.66	2928		4447
"	L3	1.69	5010		2963
pCAT CTRL	M1	0.62	32292		52379
"	M2	0.80	39821		49745
"	M3	0.68	31793		46996
ELF11	N1	1.74	9635		5523
"	N2	2.67	12969		4864
"	N3	1.48	7420		5018
	Mock	0.00	0		

SET5

construct		avg.lucif-mock	CAT-mock	(*)=exclude	CAT/LUC
(7)	a1	-0.05	588	*	
"	a2	3.68	13127		3565
"	a3	2.54	8492		3340
(6)	b1	2.34	8286		3541
"	b2	3.25	10823		3335
"	b3	2.61	9512		3648

(5)	c1	1.51	971		643
"	c2	2.39	1518		636
"	c3	2.33	1576		676
(6)	d1	2.20	6756		3074
"	d2	1.93	5540		2874
"	d3	1.62	4458		2756
(4)	e1	2.12	3249		1531
"	e2	0.66	55	*exptlerror	
"	e3	2.44	3737		1533
(5)	f1	2.35	1281		545
"	f2	1.20	950		793
"	f3	1.73	1237		716
P199/200	g1	1.62	883		544
"	g2	3.17	2156		680
"	g3	3.56	1929		542
(1)	h1	3.58	154		43
"	h2	2.92	23		8
"	h3	-0.01	-506	*	
(2)	i1	1.02	2205		2157
"	i2	2.95	2865		971
"	i3	3.50	3672		1051
(3)	j1	1.89	7414		3923
"	j2	2.70	11096		4110
"	j3	0.37	2496	*	
J-Cm	k1	1.41	527		374
"	k2	0.70	112		160
"	k3	1.89	684		362
P102/106	l1	1.58	6051		3842
"	l2	0.63	10284	*exptlerror	16455
"	l3	3.46	10659		3083
pCAT CTRL	m1	1.27	46736		36728
"	m2	1.61	49161		30630
"	m3	1.69	46945		27737
ELF11	n1	3.82	21095		5519
"	n2	3.53	19450		5514
"	n3	4.43	24321		5496
	mock1	0.00			
	mock2	0.00			

SET6

construct		avg.lucif-mock	CAT-mock	(*)=exclude	CAT/LUC
ELF11	A1	21.22	7557		356
"	A2	20.61	8400		408
"	A3	13.51	5503		407
Murine Emu	B1	33.58	2444		73
"	B2	34.26	2463		72
"	B3	41.26	2892		70
empty vec.	C1	22.16	98		4
"	C2	25.44	171		7

empty vec	C3	33.73	269	8
pCAT CTRL	D1	33.04	61750	1869
"	D2	14.37	37266	2594
"	D3	23.48	52047	2217
(1)	E1	43.34	227	5
"	E2	50.10	139	3
"	E3	21.75	40	2
(2)	F1	38.62	1777	46
"	F2	53.07	2909	55
"	F3	45.95	3309	72
(3)	G1	23.87	7614	319
"	G2	32.47	9804	302
"	G3	28.24	9522	337
P102/106	H1	50.27	13136	261
"	H2	42.55	12193	287
"	H3	29.64	8780	296
J-Cmu	I1	22.03	751	34
"	I2	19.48	573	29
"	I3	20.84	620	30
	MOCK	0.00	0	

SET7

construct		avg.lucif-mock	CAT-mock	(*)=exclude	CAT/LUC
empty vec.	A1	14.44	1228		85
"	A2	14.78	1472		100
"	A3	19.91	1787		90
"	A4	14.69	1372		93
"	A5	15.62	1423		91
"	A6	14.74	1075		73
(7)	B1	14.61	5659		387
"	B2	16.72	5709		342
"	B3	16.33	4890		299
"	B4	15.88	5489		346
"	B5	10.49	3458		330
"	B6	10.99	3249		296
(8)	C1	8.66	2147		248
"	C2	13.58	3842		283
"	C3	8.82	1754		199
"	C4	19.72	5517		280
"	C5	7.79	2232		287
"	C6	15.52	4667		301
(9)	D1	4.45	1415		318
"	D2	4.12	1145		278
"	D3	4.79	1347		281
"	D4	2.93	719		245
"	D5	6.06	1563		258
"	D6	3.94	1102		280
ELF11	E1	4.53	1605		354
"	E2	4.33	1493		345
"	E3	5.41	1800		332

ELF11	E4	5.53	1995		361
"	E5	3.07	1380		450
"	E6	4.45	1480		333
	Mock	0.00	0		

J558L

SET1

construct		avg.lucif-mock	CAT-mock	(*)=exclude	CAT/LUC
(7)	A1	0.84	4213		5039
"	A2	0.68	3209		4705
"	A3	0.48	3122		6476
(4)	E1	0.66	2484		3775
"	E2	0.60	2382		3984
"	E3	0.65	2396		3686
(5)	F1	0.56	542		965
"	F2	0.58	638		1096
"	F3	0.53	464		871
(3)	G1	0.39	2235		5790
"	G2	0.47	2294		4902
"	G3	0.57	2922		5163
pCAT CTRL	H1	0.52	3526		6730
"	H2	0.52	3440		6641
"	H3	0.62	3502		5667
pFVH-CAT-ELF11	I1	0.41	1555		3794
"	I2	0.45	1265		2811
"	I3	0.50	1539		3090
	MOCK	0.00	0		

SET2

construct		avg.lucif-mock	CAT-mock	(*)=exclude	CAT/LUC
(3)	A1	0.28	1000		3530
"	A2	0.29	1499		5229
"	A3	0.30	1073		3547
(4)	B1	0.40	653		1620
"	B2	0.41	918		2238
"	B3	0.49	782		1588
(5)	C1	0.38	61		162
"	C2	0.33	29		88
"	C3	0.34	54		159
(7)	D1	0.32	929		2934
"	D2	0.36	1218		3382
"	D3	0.48	1501		3143
ELF11	E1	0.41	795		1940
"	E2	0.37	1063		2873
"	E3	0.36	677		1908
empty vec	F1	0.57	66		115

empty vec	F2	0.60	74		122
"	F3	1.20	79		66
	MOCK	0.00	0		0

EL4

construct		avg.lucif-mock	CAT-mock	(*)=exclude	CAT/LUC
pSV-CAT-0	A1	2.28	491		216
"	A2	2.56	544		213
"	A3	2.61	439		168
pSV-CAT-P103/106	B1	5.87	890		152
"	B2	4.26	1015		239
"	B3	4.52	1815		402
pSV-CAT-SV	C1	2.55	4051		1590
"	C2	2.36	6359		2692
"	C3	2.81	5021		1786
	MOCK	0.00	0		

28S.3

construct		avg.lucif-mock	CAT-mock	(*)=exclude	CAT/LUC
pSV-CAT-0	C1	0.08	35		461
"	C2	0.10	68		667
"	C3	0.12	85		708
pSV-CAT-P103/106	D1	0.19	93		484
"	D2	0.21	116		552
"	D3	0.25	115		467
pSV-CAT-SV	E1	0.26	2052		8016
"	E2	0.31	3347		10867
"	E3	0.27	2725		10168
	mock	0.00	0		0

Appendix 4a – Table of selected high-scoring tBLASTn hits against a putative *HEB* homologue in the zebrafish genome database (version 4 at www.ensembl.org) using the inferred peptide sequence of catfish CFEB-1 (Genbank accession no. AY528668)

Query		Chromosome				Scaffold			Stats			
Start	*End*	*Name*	*Start*	*End*	*Ori*	*Name*	*Start*	*End*	*Score*	*E-val*	*%ID*	*Length*
29	199	Chr.7	53205460	53205951	+	Zv4_scaffold650	595324	595815	255	7.6e-14	35.26	190
74	145	Chr.7	50562215	50562406	-	Zv4_scaffold642	475824	476015	205	2.2e-35	56.00	75
225	271	Chr.7	50540759	50540899	-	Zv4_scaffold642	454368	454508	372	4.2e-176	89.36	47
272	319	Chr.7	50540501	50540644	-	Zv4_scaffold642	454110	454253	413	4.2e-176	95.83	48
412	479	Chr.7	50525875	50526081	-	Zv4_scaffold642	439484	439690	467	4.2e-176	81.16	69
581	636	Chr.7	50517502	50517669	-	Zv4_scaffold642	431111	431278	290	1.3e-160	67.86	56
581	658	Chr.7	50516150	50516383	-	Zv4_scaffold642	429759	429992	457	4.2e-176	76.92	78
659	705	Chr.7	50515407	50515547	-	Zv4_scaffold642	429016	429156	388	4.2e-176	91.49	47
129	181	Chr.7	50342730	50342894	-	Zv4_scaffold642	256339	256503	284	1.0e-24	61.82	55

Appendix 4b – Clustal multiple sequence alignment of catfish CFEB-1 with the predicted peptide sequence of the putative zebrafish *HEB* homologue (predicted from Ensembl transcript ID ENSDART00000009938).

```
zebrafish_predicted ------------------------------------------------------------
     catfish_CFEB-1 MNPQQRIAAIGTDKELSDLLDFSAMFSPPVNSGKNRPTTLGSSQFSASGIDERTTEAPMA

zebrafish_predicted ------------------------------------------------------------
     catfish_CFEB-1 AGGQSSPSYDESSRGFSDSPHYGDHLSDSRLVSHEGLSPTPFMSSSIMGKSERPPFSAYG

zebrafish_predicted ------------------------------------------------------------
     catfish_CFEB-1 REPGVSGCQSSLRSDMGLSSPGPVTTTGKSPAPPYSPTGSNPRRRSLQDPSPLDPLQAKK

                                                                *.****** ** ****
zebrafish_predicted ---------------------------------------------DGTHNSSDQWNLSNGI
     catfish_CFEB-1 VRKVPPGLPSSVYAPSPNSDDFNRESPSYPSSKPPSSMFASTFFDATHNSSDPWNSSNGI

                    ***.********************.**********************************:**:
zebrafish_predicted SQPGYGGMLGGSSSHMPQSGNYSNLHSHDRLNYPPHSVSPTDINASLPPMSSFHRSSAST
     catfish_CFEB-1 SQPAYGGMLGGSSSHMPQSGNYSSLHSHDRLNYPPHSVSPTDINASLPPMSSFHRSTASS

                    ********************. .: . .*.:**. :*. .:   ****:************
zebrafish_predicted SPFVTASHTPPVNTTEGVMGMLSHTRFKGGTQTNLSLSLSSTLVIYSPEHTSSSFPSNPS
     catfish_CFEB-1 SPFVTASHTPPVNTTEGVMAAANRGNATGSSQTGDALG-KALASIYSPDHTSSSFPSNPS

                    **********:.*****:**********.*********.*:.**:****.************
zebrafish_predicted TPVGSPSPLAVQAGAATAGTVVTASGPAGRAGTTQWTRTTGQAPSSPNYENSLHSLSRME
     catfish_CFEB-1 TPVGSPSPLTAQAGAASAGTVVTASGPPGRAGTTQWPRAAGQTPSSPSYENSLHSLSRME

                    ************************:*..*********:******:***:.**:******.***.*
zebrafish_predicted DRLDRLDDAIHVLRNHAVGSTAALSNDIHSLLGQAHNGPISAIGSSFPSSGLVTNRTAQM
     catfish_CFEB-1 DRLDRLDDAIHVLRNHAVGSTASLPSDIHSLLGQTHNGPITAIGTNFPTSGLVTSRTASM

                    *..****:.*** * *:.**::..*:**:********:**          .::* *******
zebrafish_predicted GSVHREESGSLNNNNHSALQASAAPTSSSELNHQADAFR--------AIAGVLASQVASP
     catfish_CFEB-1 GPAHREETASLNAN-HTGLQSTPGPASSAELNHQADTFRGSQSFALKCLSGSLASQVASA

                    * **:**********.*:***:**********:*  **.:*                   *
zebrafish_predicted LGLKVENQDKDDMHDSHASDDLKSDDESDKRDMKQNRGSSRPRQGSANYENAIHCELSCS
     catfish_CFEB-1 LELKIENQDKDDMHDNHSSDDIKSDDESDKRDIKTPRGGTR----------------TS

                    ************************************************************
zebrafish_predicted SIMEDEDLNPEQKAERERERRMANNARERLRVRDINEAFKELGRMCQLHLKSEKPQTKLL
     catfish_CFEB-1 SIMEDEDLNPEQKAERERERRMANNARERLRVRDINEAFKELGRMCQLHLKSEKPQTKLL

                    ****************************:*************:***:**:************
zebrafish_predicted ILHQAVAVILSLEQQVRERNLNPKAACIKRREEEKVSGVSGDPQQAHPAVHPGLTDTSNP
     catfish_CFEB-1 ILHQAVAVILSLEQQVRERNLNPKAACLKRREEEKVSGVSGEPQQTHPSVHPGLTDTSNP

                    ****
zebrafish_predicted HGHL
     catfish_CFEB-1 HGHL
```

Appendix 5a – Table of selected high-scoring tBLASTn hits against a putative *Oct2* homologue in the zebrafish genome database (version 4 at www.ensembl.org) using the inferred peptide sequence of catfish Oct2β (Genbank accession no. Y12652)

Query		Chromosome				Scaffold			Stats			
Start	*End*	*Name*	*Start*	*End*	*Ori*	*Name*	*Start*	*End*	*Score*	*E-val*	*%ID*	*Length*
447	466	Chr.19	28820778	28820837	+	Zv4_scaffold1635	158728	158787	148	1.8e-114	85.00	20
321	394	Chr.19	28818571	28818792	+	Zv4_scaffold1635	156521	156742	552	1.8e-114	84.00	75
243	318	Chr.19	28816413	28816640	+	Zv4_scaffold1635	154363	154590	270	1.6e-107	51.25	80
223	270	Chr.19	28816254	28816397	+	Zv4_scaffold1635	154204	154347	404	1.8e-114	100.00	48
178	223	Chr.19	28815805	28815942	+	Zv4_scaffold1635	153755	153892	257	1.8e-114	69.57	46
178	223	Chr.19	28815260	28815397	+	Zv4_scaffold1635	153210	153347	257	1.8e-114	69.57	46
73	110	Chr.19	28810493	28810600	+	Zv4_scaffold1635	148443	148550	225	1.8e-114	71.05	38
75	110	Chr.19	1085021	1085122	-	Zv4_scaffold1613	319337	319438	161	4.9e-59	63.89	36
195	223	Chr.19	1078227	1078313	-	Zv4_scaffold1613	312543	312629	237	4.9e-59	96.55	29
221	336	Chr.19	1075168	1075485	-	Zv4_scaffold1613	309484	309801	257	1.1e-35	39.69	131
305	392	Chr.19	1073181	1073444	-	Zv4_scaffold1613	307497	307760	526	4.9e-59	69.66	89

Appendix 5b – Clustal multiple sequence alignment of catfish Oct2β with the predicted peptide sequence of the putative zebrafish *Oct2* homologue (predicted from Ensembl transcript ID ENSDART00000049323).

```
zebrafish_predicted ------------------------------------------------------------
      catfish_Oct2b MTKTAPIAAMDYSHLWLPDIRMSKPIEVENPAADSPMENTGDSERNGSDSNNQIPSMKIS

                                :**:*:******:   ***:**.*:*********.***************
zebrafish_predicted ------------QAKVEDCGEMSPT--TATPAQTALTHTQLMLTGGQLAGDIQQLLQLQQ
      catfish_Oct2b PPSLSPTLSASNKAKLERCGEMSPASVQATPTQTPLSHTQLMLTGSQLAGDIQQLLQLQQ

                                ********.*.**********:*********.**************.**:*** *.**** ***
zebrafish_predicted LVLVPGHTLPSPAQFLLPQTQQQQQGLLSTPNLIPLPQQNQGSLLAAPPRLSLQAQ-REK
      catfish_Oct2b LVLVPGHPLSSPAQFLLPQAQQQQQGLLQTPNLIPLPQQNQGGLLSAPPALGLQAQQREK

                                : ***:.***.****.**************************************************
zebrafish_predicted IARSSLNTVTTVTSHPEEPSDLEELEQFARTFKQRRIKLGFTQGDVGLAMGKLYGNDFSQ
      catfish_Oct2b VMESSVSTVTPVTSHAEEPSDLEELEQFARTFKQRRIKLGFTQGDVGLAMGKLYGNDFSQ

                                **********************************************************.***
zebrafish_predicted TTISRFEALNLSFKNMCKLKPLLEKWLTDAETMSMDSTLPSPSSLSSPSLGFEGLPGRRR
      catfish_Oct2b TTISRFEALNLSFKNMCKLKPLLEKWLTDAETMSMDSTLPSPSSLSSPSLGFEGLPARRR

                                ****************:** :*** *********:*:**************************
zebrafish_predicted KKRTSIETNVRVALEHSFLANQKLTSEEILLISEKLNMEKEVIRVWFCNRRQKEKRINPS
      catfish_Oct2b KKRTSIETNVRVALERSFSTNQKPTSEEILLIAEQLNMEKEVIRVWFCNRRQKEKRINPS

                                *************:*************..*** *.* *****.********:**:*** ***
zebrafish_predicted SATPPLPSQPQPTSHKPPCYSPHMMSSQGPRQTVGSLSTAVTTMSSVCPMTPSGPSLSST
      catfish_Oct2b SATPPLPSQPQPASHKPPCYSPHMMQQQGPSQPVTSLSTAATTMSSVCPLTPTGPSSSST

                                ******.**.*:**.*.** ****.****::***************.***.****.  ***
zebrafish_predicted PSPVTPPPRGDSSPAPPTHSTLNLSTGSWHKKNGDVSNYITDFAASLSSSSQWWPAPHFQ
      catfish_Oct2b PSPVTPSPRSDASPVPATH-TLNLNTGSWRHKNGDVSNYITDFAANLSSGSQWWSNHHFQ

                                *
zebrafish_predicted S
      catfish_Oct2b S
```